OM nucleic - nucleic search, using sw model

Run on: June 24, 2004, 05:57:48 ; Search time 7590 Seconds
(without alignments)
11432.510 Million cell updates/sec

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Title: 2002
Perfect score: US-10-622-516-1
Sequence: 1 cgcctcccaagcgacctctcgg.........aaaaaaaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_vi:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_vi:*
31: em_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID	Description
1	2002	100.0	2002	6	AR399482	AR399482 Sequence
2	1712	85.5	2068	6	BC000297	Homo sapi
3	1711.2	85.5	3195	6	AK095492	Homo sapi
4	1370	68.4	1650	9	HUM3H3M	Homo sapien
5	1300	64.9	1563	9	BT007302	Homo sapi
6	1288	64.3	1685	9	X66435	H.sapiens m
7	1197.4	59.8	3260	10	BC029693	Mus muscu
8	1190.2	59.5	3202	10	BC034317	Mus muscu
9	1189.6	59.4	2703	10	BC023851	Mus muscu
10	1188	59.3	3278	10	BC031363	Mus muscu
11	1180.6	59.0	3275	6	AX700129	Sequence
12	1180.6	59.0	3275	6	AX827303	Sequence
13	1180.6	59.0	3275	10	RNCHMCA	Rattus norv
14	1103.2	55.1	127832	2	AL356361	Human DNA
c 15	1103.2	55.1	170834	2	AC021792	Homo sapi
16	1037.4	51.8	2673	9	AK025736	Homo sapi
17	798.2	39.9	1874	5	M60657	Chicken HMG
18	794.6	39.7	1824	5	I65510	Sequence 1
19	681.8	34.1	3098	5	BC042929	Xenopus 1
20	636	31.8	3205	5	BC049456	Danio rer
21	476	23.8	28001	6	AR399483	Sequence
c 22	476	23.8	144405	9	AC114947	Homo sapi
c 23	464	23.2	142657	7	AC008773	Homo sapi
24	460.4	23.0	565	6	AX320993	Sequence
25	436.4	21.8	615	6	AR225558	Sequence
26	436.4	21.8	615	6	AX321714	Sequence
27	423.4	21.1	1991	10	AK098104	Mus muscu
28	421.8	21.1	1994	10	BC024744	Mus muscu
29	421.8	21.1	2011	10	BC014714	Mus muscu
30	421.8	21.1	2098	9	BC044217	Homo sapi
31	420.2	21.0	2058	6	AX337379	Sequence
32	420.2	21.0	2058	9	AX411166	Sequence
33	420.2	21.0	2058	9	HSHMGCOAS	H.sapiens m
34	413.2	20.6	1994	6	AX163779	Sequence
35	413.2	20.6	1994	10	RATHMGCOA	Rat mitocho
36	412	20.6	476	10	HAMHMGC03	Syrian hams
37	411.4	20.5	1744	4	SSU90884	Sus scrofa
38	410	20.5	500	6	BD237076	Compounds
39	410	20.5	500	6	BD237076	Compounds
40	410	20.5	500	6	AR225476	Sequence
c 41	404	20.2	236502	2	AC098186	Rattus no
42	404	20.2	259650	2	AC107176	Rattus no
43	401	20.0	506	6	BD265188	Compounds
44	401	20.0	506	6	AR401174	Compounds
45	401	20.0	506	6	AX192620	Sequence

ALIGNMENTS

RESULT 1
AR399482
LOCUS   AR399482   2002 bp   DNA   linear   PAT 18-DEC-2003
DEFINITION Sequence 1 from patent US 6620608.
ACCESSION   AR399482
VERSION   AR399482.1 GI:40141523
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 2002)
  AUTHORS   Gong,F., Yan,C., Di Francesco,V. and Beasley,E.M.
  TITLE   Isolated human synthase proteins
  JOURNAL   Patent: US 6620608-A 1 16-SEP-2003;
FEATURES   Location/Qualifiers

```
source          1..2002
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN

Query Match         100.0%;   Score 2002;   DB 6;   Length 2002;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 2002;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
```

Qy   1 CGCCTCCCACGGGACTTCTCGGCGTGCCGGACGTGCCGGAGGTCGGGGTTGGGGGGTTATAAAGCTCGGT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CGCCTCCCACGGGACTTCTCGGCGTGCCGGACGTGCCGGAGGTCGGGGTTGGGGGGTTATAAAGCTCGGT   60

Qy   61 ACGGAAGGGGAGGGGCGGCGGGACTGTCCTTTCGTGGGGTCACTCCCTTTCCTCTGCTGCCG   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACGGAAGGGGAGGGGCGGCGGGACTGTCCTTTCGTGGGGTCACTCCCTTTCCTCTGCTGCCG   120

[DNA sequence alignment continues, Query (Qy) and Database (Db) rows with 100% identity, positions 1 through 2002]

Qy   1981 AAAAAAAAAAAAAAAAAAAA   2002
           ||||||||||||||||||||
Db   1981 AAAAAAAAAAAAAAAAAAAA   2002

```
RESULT 2
LOCUS       BC000297              2068 bp    mRNA    linear   PRI 30-SEP-2003
DEFINITION  Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
            (soluble), mRNA (cDNA clone IMAGE:2819708), partial cds.
ACCESSION   BC000297
VERSION     BC000297.2  GI:33991030
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2068)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22388257
   PUBMED    12477932
REFERENCE   2  (bases 1 to 2068)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     On Aug 20, 2003 this sequence version replaced gi:12653064.
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 1 Row: a Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4504428.
FEATURES             Location/Qualifiers
     source          1..2068
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
RESULT 2
LOCUS       BC000297              2068 bp    mRNA    linear   PRI 30-SEP-2003
DEFINITION  Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
            (soluble), mRNA (cDNA clone IMAGE:2819708), partial cds.
ACCESSION   BC000297
VERSION     BC000297.2  GI:33991030
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2068)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22388257
   PUBMED    12477932
REFERENCE   2  (bases 1 to 2068)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
COMMENT     On Aug 20, 2003 this sequence version replaced gi:12653064.
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
  REMARK    Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 1 Row: a Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4504428.
FEATURES             Location/Qualifiers
     source          1..2068
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2819708"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..2068
                     /gene="HMGCS1"
                     /note="synonym: HMGCS"
                     /db_xref="LocusID:3157"
                     /db_xref="MIM:142940"
     CDS             <1..1668
                     /gene="HMGCS1"
                     /codon_start=1
                     /product="HMGCS1 protein"
                     /protein_id="AAH00297.2"
                     /db_xref="GI:33991031"
                     /db_xref="LocusID:3157"
                     /db_xref="MIM:142940"
                     /translation="HSLSSAAARSRLCPKEEIVTDLETAVLYPSHSSFTMPGSLPLNA
                     EACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINS
                     LCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTIEGIDT
                     TNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAYYATGNARPTGGVGAVALLIGP
                     NAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVVCKKIH
                     AQWQKENDKDITLNDGFMIFHSPYCKLVQKSLARMLLNDFLNDQRDKNSIYSGLE
                     AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTASLLVSNQNGNMYTSSVGSLASVL
                     AQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTDDTPGSALDKITASLCDLKSRLDSR
                     DDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH"
     misc_feature    142..1512
                     /gene="HMGCS1"
                     /note="HMG CoA synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CDD:pfam01154"
ORIGIN
        Query Match     85.5%;  Score 1712;  DB 9;  Length 2068;
        Best Local Similarity   93.6%;
        Matches 1848;  Conservative    0;  Mismatches    0;  Indels  126;  Gaps    1;

Qy        134 GCTCTTTCACCATGGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         95 GCTCTTTCACCATGGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 154

Qy        194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        155 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 214

Qy        254 AAAAAATATGATACGGTCGTAGAGTCGGAAAAGTATACCATTGGCTTGGGCGCCAGGGATGG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        215 AAAAAATATGATACGGTCGTAGAGTCGGAAAAGTATACCATTGGCTTGGGCGCCAGGATGG 274

Qy        314 GCTTCTGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        275 GCTTCTGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 334

Qy        374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGCGGCTGGAAGTTGGAACAGAGACAA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        335 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGCGGCTGGAAGTTGGAACAGAGACAA 394

Qy        434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGGTCGGGA 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        395 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGGTCGGGA 454

Qy        494 ATACAGATATAGAAGGAATCGACACAACTAATGCATATGGAGGCACAGCTGCTGTGTTCT 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        455 ATACAGATATAGAAGGAATCGACACAACTAATGCATATGGAGGCACAGCTGCTGTGTTCT 514

Qy        554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATG-------------------- 591
              |||||||||||||||||||||||||||||||||||||||
Db        515 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 574

Qy        592 ---------- 591
Db        515 ---------- 591
```

```
Qy    575 CAGGAGATATTGCTGTATATGCCACAGGAAATGCTAGAGCCTACAGGTGGGAGTTGGAGCAGAG 634
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1655 ATGGGGGAACATTAAGATACTTCTGTGAGGGTGCAAGACGACTTCAGGGTGGGGTCGGGGT 1714

Qy    592 -------------------------------------------GGGCTTGGTGGCACAC 607
                                                     ||||||||||||||||||
Db   1628 GGGGGTATGGGAACACAGTTGGAGGAATGGGGATATCTCGGGGGGGATAATTTTTAAAGGATTACATG 1687

Db   1715 GGGGGGTATGGGAACACAGTTGGAGGAATGGGGATATCTCGGGGGGGGATAATTTTTAAAGGATTACATG 1774

Qy    635 TAGCTCTGCTAATTGGGGCCAAATGCTCCTTTAATTGGAACGGAGGGCCTTCGTGTGGGACAC 694
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1688 TTATGTAAAATTTTTATGTGACTGAACATGAGAGCCCTGGAGTGACATCAGTGTACTTGGGAAAG 1747

Qy    608 ATATGCAACATGCCTATGATTTTTTACAAGCCTGAATAGTGCTATCGTGAATATCCTATAGTAG 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1775 TTATGTAAAATTTTTATGTGACTGAACATGAGAGCCCTGGAGTGACATCAGTGTACTTGGGAAAG 1834

Qy    695 ATATGCAACATGCCTATGATTTTTTACAAGCCTGAATAGTGCTATCGTGAATATCCTATAGTAG 754
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1748 TCTCTTTGCTCTATTTTGCTGACATGCTTCCTTGTTTGGTGTCTGGCCAATGCCAAAATGTACT 1807

Qy    668 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGGACCGCTGCTACTTCACTTGTCTTACTT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1835 TCTCTTTGCTCTATTTTGCTGACATGCTTCCTTGTTTGGTGTCTGGCCAATGCCAAAATGTACT 1894

Qy    755 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGGACCGCTGCTACTTCACTTGTCTACT 814
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1808 CGAATGATGTTAAGGGGCTCTGTAAAACTTCATCACCTTCTTGGTGGTTCGGCCATTTGTATGCAGTCATGAT 1867

Qy    728 GCAAAAAGATCCATGCCCCAGTGGGCACAGAAAAGAGGGGAAAAATGATAAAGATTTACCTTGAATG 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1895 CGAATGATGTTAAGGGGCTCTGTAAAACTTCATCACCTTCTTGGTGGTTCGGCCATTTGTATGCAGTCATGAT 1954

Qy    815 GCAAAAAGATCCATGCCCCAGTGGGCACAGAAAAGAGGGGAAAAATGATAAAGATTTACCTTGAATG 874
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1868 TTTGGTTTTTAAACATGGTATAATGATGAATTTGTGTACTTCTGTCAGAAGAAGCACAGGAGGTAC 1927

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 TTTGGTTTTTAAACATGGTATAATGATGAATTTGTGTACTTCTGTCAGAAGAAGCACAGGAGGTAC 2014

Qy    875 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 934
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1928 TAACTCCAATTAAAAAAATTTTTTAACATGTTAAAAAAAAAAAAAAAAAAAAAAAAA 1981

Qy    935 GGAATGTTGCTGAATGACTTCCTTAATGACACAGAATAGAGGATAAAAATAGTAATCTATAGTG 994
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1955 TTTGGTTTTTAAACATGGTATAATGATGAATTTGTGTACTTCTGTCAGAAGAAGCACAGGAGGTAC 2014

Db   1928 TAACTCCAATTAAAAAAATTTTTTAACATGTTAAAAAAAAAAAAAAAAAAAAAAAAA 1981

Qy    908 GCCTGGAAGCCTTGGGGATGGTTAAATTAGAAGACACCTACTTTTGATTAGAGATTGGGAGA 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1928 TAACTCCAATTAAAAAAATTTTTTAACATGTTAAAAAAAAAAAAAAAAAAAAAAAAA 1981

Qy    995 GCCTGGAAGCCTTTGGGGATGGTTAAATTAGAAGACACCTACTTTGATTAGAGAGTTGGGAGA 1054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2015 TAATCTCCAATTAAAAAATTTTTTTTAACATGTGTAAAAAAAAAAAAAAAAAAAAAA 2068

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCCAGTGAAAAACAAAGGCATTTTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1055 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCCAGTGAAAAACAAAGGCATTTTTACTTG 1114

Qy   1028 TATCAAATCAAAAATGGAAAATATGTACACATATCTTCAGTATATGGTTCCTTGCATCTGTTC 1087
Qy   1115 TATCAAATCAAAAATGGAAAATATGTACACATATCTTCAGTATATGGTTCCTTGCATCTGTTC 1174

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCACGAGGGGAAGAGAATTGGGAGTGTTTTTCTTATGGTT 1147
Qy   1175 TAGCACAGTACTCACCTCAGCAATTAGCACGAGGGGAAGAGAATTGGGAGTGTTTTTCTTATGGTT 1234

Qy   1148 CTGGTTTGGCTGCACACTCTGTACTCTCTAAAGTCACCAAGATGCTACACACCGGGTCTCTG 1207
Qy   1235 CTGGTTTGGCTGCACACTCTGTACTCTCTAAAGTCACCAAGATGCTACACACCGGGTCTCTG 1294

Qy   1208 CTCCTTGATAAAAATAACACAGCAAGTTATGGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
Qy   1295 CTCTTGATAAAAATAACACAGCAAGTTATGGATCTTAAATCAAGGCTTGATTCAAGAACTG 1354

Qy   1268 GTGTGGCACCAGATGCTTCTTCGGCTGAAAAACATGAAGCTTCAGAGGACACCCATCATTTGG 1327
Qy   1355 GTGTGGCACCAGATGCTTCTTCGGCTGAAAAACATGAAGCTTCAGAGGACACCCATCATTTGG 1414

Qy   1328 TCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTGNAAGGAAACGTGGTACTTAGTTA 1387
Qy   1415 TCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTGNAAGGAAACGTGGTACTTAGTTA 1474

Qy   1388 GGGTGGATGAAAAGCACAGAGAACTTACGCTGCTGGCCGTCCCACTCCAAATGATGACACTT 1447
Qy   1475 GGGTGGATGAAAAGCACAGAGAACTTACGCTGCTGGCCGTCCCACTCCAAATGATGACACTT 1534

Qy   1448 TGGATGAAGAGGAGTAGGACTTGTGTCGATTCAAATCAATGAGCATATTCCAAGCCCTG 1507
Qy   1535 TGGATGAAGAGGAGTAGGACTTGTGTCGATTCAAATCAATGAGCATATTCCAAGCCCTG 1594

Qy   1508 CCAAGAAAGTACCAAGACACTCCCTCCCACACAGCAGCAGAACTTGAACGCGTGTTCATTAGTTA 1567
Qy   1595 CCAAGAAAGTACCAAGACACTCCCTCCCACACAGCAGCAGAACTTGAACGCGTGTTCATTAGTTA 1654

Qy   1568 ATGGGGGAACAATTAAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTCGGGGCATGGGGT 1627
Db   ...
```

RESULT 3
AK095492

LOCUS   AK095492   3195 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION   Homo sapiens cDNA FLJ38173 fis, clone FCBBF1000053, highly similar
               to HYDROXYMETHYLGLUTARYL-COA SYNTHASE, CYTOPLASMIC (EC 4.1.3.5).
ACCESSION   AK095492
VERSION   AK095492.1 GI:21754757
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Oshima,A., Takahashi-Fujii,A., Tanase,T., Imose,N., Takeuchi,K.,
               Arita,M., Musashino,K., Yuuki,H., Hara,H., Sugiyama,T., Irie,R.,
               Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
               Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
               Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
               Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Sugiyama,A.,
               Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
               and Isogai,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 3195)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
               Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
               (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
               Economy, Trade and Industry of Japan; cDNA full insert sequencing:
               Research Association for Biotechnology (RAB); cDNA library
               construction: Helix Research Institute (HRI) (supported by Japan
               Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
               HRI, and Biotechnology Center, National Institute of Technology and
               Evaluation; clone selection for full insert sequencing: HRI and
               RAB; annotation: HRI and RAB.
FEATURES   Location/Qualifiers
     source   1..3195
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="FCBBF1000053"
                     /tissue_type="brain"
                     /clone_lib="FCBBF1"
                     /dev_stage="fetal"

```
CDS    68..1597
       /note="cloning vector: pMl8SFL3"
       /note="unnamed protein product"
       /codon_start=1
       /protein_id="BAC04559.1"
       /db_xref="GI:21754758"
       /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
       GKYTIGLGQARMGFCPTDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
       SVKTNLMQLFEESGNTDIEGIDTTNAVNWIESSSWDGRYALVAGDIAVVARGNARPT
       GGVGAVALLIGPNAPLIFERGLRGTHMQHAVDFYKPDMLSEYPIVDGKLSILCYLSAL
       DRCYSVVCKKIHAQWQKEGNDKDFTLNDFGFRMIFHSPYCKLVQKSLARMLLNDFLNDQ
       NRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAPMKASSELFSQKTKASLLVSNQNGNMY
       TSSVGSLASLVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTDPATPSSALDKITA
       SLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNVIPQGSIDSLFEGTWLIVRDVEK
       HRRTYARRPTENDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPERAAVISNGE
       H"
ORIGIN
```

```
Qy     78  GCGGACTGTCCTTTCGTTCGGCCTCACTCCCTCCCCCTTCGGTCCTCGGTCACCGCTTGCTC  137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GCGGACTGTCCTTTCGTTCGGCCTCACTCCCTCCCCCTTCGGTCCTCGGTCACCGCTTGCTC  60

Qy    138  TTTCACCATGCCTGGGATCACTTCCTTTGAAATGCAGAAGCTTGCTGCCCCTCGGTCGGTGG  197
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TTTCACCATGCCTGGGATCACTTCCTTTGAAATGCAGAAGCTTGCTGCCCCTCGGTCGGTGG  120

Qy    198  AATTGTTGCCCTTGAGACTATATTTCCTTCTCATATATGTTGATCAAGCAGAGTTGGAAAA  257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AATTGTTGCCCTTGAGACTATATTTCCTTCTCATATATGTTGATCAAGCAGAGTTGGAAAA  180

Qy    258  ATATGATGGTGTAGAGTGCTGGAAAGTATAACCATTGGGCTTGGCCCCAGGCCAAGAGAGT  317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ATATGATGGTGTAGAGTGCTGGAAAGTATAACCATTGGGCTTGGCCCCAGGCCAAGAGAGT  240

Qy    318  CTGCACAGATAGAGAAGAGATATTAACTCTCTTGCTTTGCATGACTGTGGTTCAGATCTTA  377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CTGCACAGATAGAGAAGAGATATTAACTCTCTTGCTTTGCATGACTGTGGTTCAGATCTTA  300

Qy    378  GAGAAATAACCTTTCCTATGATTGCATGGGCGGCGCTGGAAAGTTGGAACAGACAATCAT  437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GAGAAATAACCTTTCCTATGATTGCATGGGCGGCGCTGGAAAGTTGGAACAGACAATCAT  360

Qy    438  CGACAAATCAAAGTCTGTGAAGACTAATTGATGGAGTCTGGGAATAC  497
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CGACAAATCAAAGTCTGTGAAGACTAATTGATGGAGTCTGGGAATAC  420

Qy    498  AGATATAGAAGGAATCGACACAACTAATCGC----------  527
           ||||||||||||||||||||||||||||||||
Db    421  AGATATAGAAGGAATCGACACAACTAATCGC----------  480

Qy    528  ---ATGCTATGGAGGCACAGCTCGTCTTCTTCAATGCTGTTAACTGGA---------  571
              |||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGGACGGTATGCCCTGGGTAGTTGCAGGGAGAATAATTGCTGTTAACTGGA  540

Qy    572  -------------TTGAGTCCAGCTC  584
                        ||||||||||||
Db    541  TCCACCAGGAAATGCTAG  600
```

[The sequence alignment continues with additional Qy/Db paired rows extending through to positions 721/780 on the left portion and positions 765/824 through 1801/1860 on the right portion of the page.]

```
Qy    541  ACCTACAGGTGGGACGTAGCTCTGCTAATTGGGCAAATGCTTCCTTTACTT  600
Db    601  TGAACGAGGGCCTTCGTGGGACACATATGCACCTATGATTTTTACAAGCC  660
Qy    601  TGAACGAGGGCCTTCGTGGGACACATATGCACCTATGATTTTTACAAGCCTGATAT  660
Db    661  GCTATCTGAATATCCTATAGTAGTAGAAACACTCTCCAATGCCTATGAAGTCTC  720
Qy    645  GCTATCTGAATATCCTATAGTAGTAGAAACACTCTCCAATGCCTATGAAGTCTCCTTCAGTGTGCATT  704
Db    721  AGACCGCTGCTATTCTGTCTACTCGCAAAAAGATCCATCCCCAGTGGCAGAAAGAGGGAAA  780
Qy    705  AGACCGCTGCTATTCTGTCTACTCGCAAAAAGATCCATCCCCAGTGGCAGAAAGAGGGAAA  764
Db    785  GTCTCTGGCCAAATGCCAAATGTACTCGAAATGAGTTAAGGGCTCTGTAAAAACTTCATACCTC  844
Qy    721  AGACCGCTGCTATTCTGTCTACTCGCAAAAAGATCCATCCCCAGTGGCAGAAAGAGGGAAA  780
```

```
Qy   1845 TTTGGGCCATTTGTATGCCATCGATGTTTGGTTTTAAACATGGTAATGAATGTTGTGTACTT 1904
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TTTGGGCCATTTGTATGCCATCGATGTTTGGTTTTAAACATGGTAATGAATGTTGTGTACTT 1920

Qy   1905 CTGTCAGAAGAGAAAGCAGAGGTACTAATCTCCAATTAAAATAAATTTTTAACATGTAAAA 1964
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGTCAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAATAAATTTTTAACATGTAAGAAT 1980

Qy   1965 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1994
          |||||||||||||||||||||||||||||||||
Db   1981 TTTGTACTTTGAACAACAAGATTACAGAAA 2010
```

RESULT 4
HUM3H3M

LOCUS	HUM3H3M	1650 bp mRNA linear PRI 24-FEB-1995

DEFINITION Homo sapiens 3-hydroxy-3-methylglutaryl coenzyme A synthase mRNA,
            complete cds.
ACCESSION   L25798
VERSION   L25798.1 GI:410027
KEYWORDS   3-hydroxy-3-methylglutaryl coenzyme A synthase.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1650)
  AUTHORS   Rokosz,L.L., Boulton,D.A., Butkiewicz,E.A., Sanyal,G., Cueto,M.A.,
            Lachance,P.A. and Hermes,J.D.
  TITLE   Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase:
            expression, purification, and characterization of recombinant
            wild-type and Cys129 mutant enzymes
  JOURNAL   Arch. Biochem. Biophys. 312 (1), 1-13 (1994)
  MEDLINE   94304197
   PUBMED   7913309
COMMENT   Original source text: Homo sapiens fetal adrenal cDNA to mRNA.
FEATURES   Location/Qualifiers
     source   1..1650
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="fetal adrenal"
     CDS   22..1584
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl coenzyme A synthase"
                     /protein_id="AAA62411.1"
                     /db_xref="GI:410028"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDA
                     GKYTIGLGQAKWGFCTDRDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
                     SVKTNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
                     AVYATGNARPTGGVGAVALLIGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDG
                     KLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGEMIFHSPYCKLVQKSLA
                     RMLLNDFLNDKDRKQNSIYSGLEAPGDVKLEDTYFDRDVEKAFMKASSELFSQKTKAS
                     LLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTGVAPVDFAENMKLREDTHHLVNYIPQGSIDSLFE
                     GTWYLVRVDEKHRRTYARRPTPNDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAA
                     EPEAAVISNGEH"
ORIGIN

Query Match   68.4%; Score 1370; DB 9; Length 1650;
Best Local Similarity   92.0%; Pred. No. 1e-287;
Matches 1509; Conservative   0; Mismatches   5; Indels 126; Gaps   1;
```

```
Qy    134 GCTCTTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTTGGCCAAAAGATG 193
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 GCTCTTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTTGGCCAAAAGATG 70

Qy    194 TGGGAATTGTTGCCCTTGAGATCTATTTCCTTCCTTTCAATATGTTGATCAAGCAGAGTTGG 253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TGGGAATTGTTGCCCTTGAGATCTATTTCCTTCCTTTCAATATGTTGATCAAGCAGAGTTGG 130

Qy    254 AAAAATATGATGGGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 AAAAATATGATGGGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 190

Qy    314 GCTTCTGCACAGATAGAGAAGAAGAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 GCTTCTGCACAGATAGAGAAGAAGAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 250

Qy    374 TGGAGAGAAAATAACCTTTCCTTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAAGAACAA 433
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 TGGAGAGAAAATAACCTTTCCTTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAAGAACAA 310

Qy    434 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGTGTGTTTGATGGAGAGTCTGGGA 493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGTGTGTTTGATGGAGAGTCTGGGA 370

Qy    494 ATACAGATATAGAAGGAGAATCGACACAACTAATGCATGCTATGCTGGAGGCAACAGCTGCTGTGTCT 553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 ATACAGATATAGAAGGAGAATCGACACAACTAATGCATGCTATGCTGGAGGCAACAGCTGCTGTGTCT 430

Qy    554 TCAATGCTGTGTTAACTGATTGAGTCCAGCTCTTGGGGAT----------------------- 591
          |||||||||||||||||||||||||||||||||||||||||
Db    431 TCAATGCTGTGTTAACTGATTGAGTCCAGCTCTTGGGGAT------------------------- 490

Qy    592 --------------------------------------------------------------- 591
Db    491 CAGGAGAGATATTGCGTGTATATGCCACAGGGAAAATGCCTAGACCTACAGGGTGGAGTTGGAGCAG 550

Qy    592 ---------------------------------GGGCTTCGTCGGGACAC 607
                                            |||||||||||||||||
Db    551 TAGCTCGTGCTTAATTGGGCCAAATGCTCCTTTAATTTTGGAACGAGGCGCTTGGTGGGACAC 610

Qy    608 ATATCGACACATGCCTATGCATTTTTTACCAAGGCCTGATATGGCTATCTGAATATCCTATAGTAG 667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATCGACACATGCCTATGCATTTTTTACCAAGGCCTGATATGGCTATCTGAATATCCTATAGTAG 670

Qy    668 ATGGAAAAACTCTCCATACAGTGGCTACCTCAGTGCATTAGACCGCGTCGTACTCTGTCTACT 727
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGAAAAACTCTCCATACAGTGGCTACCTCAGTGCATTAGACCGCGTCGTACTCTGTCTACT 730

Qy    728 GCAAAAAGATCCAGTGGCACAAAGAGAGGGAAAATGATAAAGATTTTACCTTGAATG 787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCAGTGGCACAAAGAGAGGGAAAATGATAAAGATTTTACCTTGAATG 790

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAAATCTCTAGCTC 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAAATCTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATGACCAGATAGAGATAAAAATAGTATGTATCTATAGTG 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGATAGAGATAAAAATAGTATGTATCTATAGTG 910

Qy    908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGA 967
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGA 970

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTACTTG 1030

Qy   1028 TATCAAATCAAAAATGAAAATAGTACACATCTTTCAGTAGTATGGTTCCCATCGTTC 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAATCAAAAATGAAAATAGTACACATCTTTCAGTAGTATGGTTCCCATCGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAGAGTTTTCTTATGGTT 1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAGAGTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGGCTGCCACTCTTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGGCTGCCACTCTTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1210

Qy   1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1270
```

```
Qy   1268  GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTGG  1327
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271  GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTGG  1330

Qy   1328  TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA  1387
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331  TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA  1390

Qy   1388  GGGTGGATGGAAAGCACAGAAGAACTTAGCTCGGCGTCCCACTCCAAATGATGACACTT  1447
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391  GGGTGGATGGAAAGCACAGAAGAACTTAGCTCGGCGTCCCACTCCAAATGATGACACTT  1450

Qy   1448  TGGATGAAGGAGTGAGGAGTAGGACTTGTGCCATTCAAACATAGCAACATTCCAAGCCCTG  1507
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451  TGGATGAAGGAGTGAGGAGTAGGACTTGTGCCATTCAAACATAGCAACATTCCAAGCCCTG  1510

Qy   1508  CCAAGAAAGTACCAACAGACTCCCCTGCCACAGCACCAGCAGCTGTGTCATTAGTA  1567
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511  CCAAGAAAGTACCAACAGACTCCCCTGCCACAGCACCAGCAGCTGTGTCATTAGTA  1570

Qy   1568  ATGGGGAACATTAAGATAACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT  1627
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571  ATGGGGAACATTAAGATAACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT  1630

Qy   1628  GGGGGTATGGGAACAGTTGG  1647
           ||||||||||||||||||||
Db   1631  GGGGGTATGGGAACAGTTGG  1650
```

```
RESULT 5
BT007302
LOCUS       BT007302               1563 bp    mRNA            linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
            (soluble) mRNA, complete cds.
ACCESSION   BT007302
VERSION     BT007302.1  GI:30583442
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1563)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1563)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1..1563
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH008878X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
```

```
                     /lab_host="DH5alpha T1 resistant"
                     /note="vector: pDNR-Dual"
     CDS             1..1563
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
                     (soluble)"
                     /protein_id="AAP35966.1"
                     /db_xref="GI:30583443"
                     /translation="MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQRELEKIDGVDA
                     GKYTIGLGQAKMGFCTTREDINSLCMTVVQNLMERNNLSYDIGRLEVGTEIIIDKSK
                     SVKTNLMQLFEESGNTDLEGIDTTNACKDGVGAVFNAVNWIESSSWDGRYALVVAGDI
                     AVYATGNARPTGGVGAVALLIGPNAPLIFERGLRGTHMQHAYDPYKPDMLSEYPIVDG
                     KLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMLLNDFLNDQRDKNSIYSGLEAFGDVKLEDYFDRDVEKAFMKASSELFSQKTKAS
                     LIVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFE
                     GTWYLVRVDEKHRRTYARRPTNDDTLDEGVLVHSNIATEHIPSPAKKVPRLPATAA
                     EPEAAVISNGEH"
ORIGIN
```

```
Query Match       64.9%;  Score 1300;  DB 9;  Length 1563;
Best Local Similarity   91.9%;  Pred. No. 1.7e-272;
Matches 1436;  Conservative  0;  Mismatches  0;  Indels 126;  Gaps 1;

Qy   145  ATGGCCTGGATCAATCACTTCCTTTGAATGCAGAAGCTTGCTCGGCCCAAAAGATGTTGTT  204
          |||||||||||| ||||||||||||||||||||||||| ||||||||||||||||||  ||
Db     1  ATGGCCTGGATCACTTCCTTTGAATGCAGAAGCTTCCTCGGCCCAAAAGATGTGGGAATTGTT  60

Qy   205  GCCCTTGAGATCTATTTTCCTTCTCAAATAATGTTGATCAAGCAGAGTTGGAAAAATATGAT  264
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  GCCCTTGAGATCTATTTTCCTTCTCAAATAATGTTGATCAAGCAGAGTTGGAAAAATATGAT  120

Qy   265  GGTGTAGATCTGGAAAGTATAACCATTGGCTTGGGCCAGGCCAAGAGCTTCTGCACA  324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GGTGTAGATCTGGAAAGTATAACCATTGGCTTGGGCCAGGCCAAGAGCTTCTGCACA  180

Qy   325  GATAGAGAAGATATTAAACTCTCTTGCACTGACTGTGGTTCAGATTCTTATGGAGAGAAT  384
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  GATAGAGAAGATATTAAACTCTCTTGCACTGACTGTGGTTCAGATTCTTATGGAGAGAAT  240

Qy   385  AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGAGACAATCATCGACAA  444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGAGACAATCATCGACAA  300

Qy   445  TCAAAGTCTCGTGAGACTAATTTGATGCACTGTTTGAAGAGTCTGGGAATACAGATATA  504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  TCAAAGTCTCGTGAGACTAATTTGATGCACTGTTTGAAGAGTCTGGGAATACAGATATA  360

Qy   505  GAAGGAAATCGACACAACTAATGCATGCTATGGAGGCACAGCCTGCTGTGTTCTCAAATCTGTT  564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GAAGGAAATCGACACAACTAATGCATGCTATGGAGGCACAGCCTGCTGTGTTCTCAAATCTGTT  420

Qy   565  AACTGGATTGAGTCCAGCTCTTGGGAT  591
          |||||||||||||||||||||||||||
Db   421  AACTGGATTGAGTCCAGCTCTTGGGATATGCCCTGGTAGTTGCAGGAGATATT  480

Qy   592  -----------------------------------------------------  591
          
Db   481  GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTGCTA  540

Qy   592  -----------------------------------------------------  591
          
Db   541  ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACACATAT  600

Qy   592  -----------------------------------------------------  591
          
Db   601  GCCTTATGCATTTTACAAGCCTGATAATGCTATCTGAATAATCCTATATAGTGAGAAAAACTC  660

Qy   592  -----------------------------------------------------  591
          
Db   661  TCCATACAGTGCTACCTCAGTGCATTAGACCGGCTTCGTCTACTGGAGAGATC  720

Qy   592  -----------------------------------------------------  591
          
Db   721  CATGCCCAGTGGCGAAAAGAGGGGAAATGATAAAGATTTACCTTGAATGATTTTGGCTTC  798
```

RESULT 6

```
LOCUS       HSCOAS       1685 bp    mRNA    linear   PRI 12-SEP-1993
DEFINITION  H.sapiens mRNA for HMG-CoA-synthase.
ACCESSION   X66435
VERSION     X66435.1  GI:30008
KEYWORDS    Hydroxymethylglutaryl CoA Synthase.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1685)
AUTHORS     Russ,A.P., Ruzicka,V., Maerz,W., Appelhans,H. and Gross,W.
TITLE       Amplification and direct sequencing of a cDNA encoding human
            cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase
JOURNAL     Biochim. Biophys. Acta 1132 (3), 329-331 (1992)
MEDLINE     93041999
PUBMED      1358203
REFERENCE   2  (bases 1 to 1679)
AUTHORS     Russ,A.
TITLE       Direct Submission
JOURNAL     Submitted (26-MAY-1992) A. Russ, Labor fur angewandte Biochemie,
            Theodor Stern-Kai 7, W-6000 Frankfurt am Main 70, FRG
REMARK      revised by [2]
REFERENCE   3  (bases 1 to 1685)
AUTHORS     Russ,A.
TITLE       Direct Submission
JOURNAL     Submitted (10-AUG-1992) Andreas Russ, Zentrum der biologischen
            Chemie, J.W.-Goethe-Universitaet Frankfurt, Theodor-Stern-Kai 7,
            Frankfurt, 6000, Germany
FEATURES             Location/Qualifiers
     source          1..1685
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HMGCoASynthase"
                     /cell_type="fibroblast"
     CDS             123..1685
                     /codon_start=1
                     /product="hydroxymethylglutaryl CoA Synthase"
                     /protein_id="CAA47061.1"
                     /db_xref="GI:30009"
                     /db_xref="GOA:Q01581"
                     /db_xref="SWISS-PROT:Q01581"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                     GKYTIGLGQAKMGFCTDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
                     SVKTNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
                     AVYATGNARPTGGVGAVALLIGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDG
                     KLSIQCYLSALDRCYSVYCKKIHAQWQKEANDNFFILNDPGFMIFHSPYCKLVQKSLA
                     RMLLNDFLNDQNRDKNSIYSGLKAFGDVKLIEDTYFDRDVEKAFMKASSELFSQKTKAS
                     LLVSNQNGNMYTSSVYGSLASVLAQYSPQHLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTGVAQDVFAENMKLREDTHHLVNYIPQGSIDSLFE
                     GTWYLVRVDEKHRRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAA
                     EPEAAVISNGVW"
ORIGIN
```

Query Match   64.3%;   Score 1288;   DB 9;   Length 1685;
Best Local Similarity   91.0%;
Matches 1433;   Conservative   0;   Mismatches 15;   Indels 126;   Gaps 1;
Pred. No. 7.1e-270;

```
Qy   134 GCTTCTTTCACCATGCCTGGATCACTTCCTTTGAATCCAGAAGCTTGCTCGGCCAAAAGATG 193
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   112 GCTCTTTCACCCATGCCTGGATCACTTCCCTTGAATGCAGAAGCTTGCTCGGCCAAAAGATG 171

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAAGCAGAGTTGG 253
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAAGCAGAGTTGG 231

Qy   254 AAAAATATGATTGGTCTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGAATGG 313
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 AAAAATATGATTGGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGAATGG 291

Qy   314 GCTTCTGCACAGATAGGAAGAGATATTAACTCTCTTTGCATGACATGTGGTTGGTTCAGAAT CTTA 373
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   292 GCTTCTGCACAGATAGGAAGAGATATTAACTCTCTTTGCATGACATGTGGTTGGTTCAGAATCTTA 351

Qy   374 TGGGAGAGAAATAACCTTTCCTATGATTGCATTGCATTGGGCGGCGGCCTGGAACACAGAGACA 433
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   352 TGGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGCCTGGAACACAGAGACA 411

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGA 493
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   412 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGA 471

Qy   494 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGACACAGCTGCTGTCT 553
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   472 
```

```
Qy   721 CATGCCCAGTGGCAGAAGAGGCAGAAATGATAAAGAGATTTTACCTTGAATGATTTGGCTTC 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ... 

Qy   799 ATGAGACTTCACTCACCCATATGTGAAATCTTCAGAAATCTTAGCTCGGATGTGTGCTG 858
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ATGATCTTTCACTCACCATATTGTAAATCTGGTTCAGAATCTCTAGCCGATGTGTGCTG 840

Qy   859 AATGACTTCCTTAATCAGACCAGAATACAGATAAAAATCTATATTATAGTGGCCTGGAAGCC 918
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 AATGACTTCCTTAATGACCAGAATAGGAATACAGATAAAAATAGTATCTATATGTGGCCTGGAAGCC 900

Qy   919 TTTGGGGATGTTAAATTAGAGACAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 978
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TTTGGGGATGTTAAATTAGAGACAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 960

Qy   979 AAGGCTAGCTTCTGAAACTCTTCAGTCAGCAGAAACAAAAGGCATCTTTACTTGATTCAAAATCAA 1038
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   ...

Qy   961 AAGGCTAGCTTCTGAAACTCTTCAGTCAGCAGAAACAAAAGGCATCTTTACTTGATTCAAAATCAA 1020

Qy  1039 AATGGAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCATCTGTTCTTCTAGCACAGTAC 1098
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 AATGGAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCATCTGTTCTTCTAGCACAGTAC 1080

Qy  1099 TCACCTCAGCAATTAGCAGGGAAGAGAAAGAAATTGGAGTGTTTTTGTTTCTTATGGTTTCTGGTTCTGGT 1158
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TCACCTCAGCAATTAGCAGGGAAGAGAAAGAAATTGGAGTGTTTTTGTTTCTTATGGTTTCTGGTTTGGCT 1140

Qy  1159 GCCACTCTGTACTCTCTTTAAAGTCACACAGATGCTACACGGGGTCTGCTCTTGATAAA 1218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACGGGGTCTGCTCTTTGATAAA 1200

Qy  1219 ATAACAGCAAAGTTTATGTGATCTTAAAATCAAAGGCTTCAGAGAACTGGTGTTGGGCACCA 1278
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATAACAGCAAAGTTTATGTGATCTTAAAATCAAAGGCTTCAGAGAACTGGTGTTGGCACCA 1260

Qy  1279 GATGTTCTTCGCTGAAAACATGAAGCTTCAGAGAGAGCTTCAAACTATATT 1338
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GATGTTCTTCGGCTGAAAAACATGAAGCTTCAGAGAGAGGACACCCCATCATCAGTTGGTCAACTATATT 1320

Qy  1321 CCCAGGGGTTCAATAGAATTCACTCTTTTGAAGGACAGTGGTACTTAGTTAGGGTGGATGAA 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 CCCAGGGGTTCAATAGAATTCACTCTTTGAAGGACAGTGGTACTTAGTTAGGGTGGATGAA 1380

Qy  1381 AAGCACGAGAGAAGAACTTAGGCTCGGCGGCGTCCCACTCCAAAATGATGACACTTTGGATGAAGGA 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  ...

Qy  1399 AAGCACGAGAGAAGAACTTAGGCTCGGCGGCGTCCCACTCCAAAATGATGACACTTTGGATGAAGGA 1458

Qy  1459 GTAGGACTTGTTGCATTCAAACAATTAGCACAATCTAAAATTCCAAGCCCTGCCAAGAAAGTA 1518
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 GTAGGACTTGTTGCATTCAAACAATTAGCAACTAGCAACTGGAGCATTAATTCCAAGCCCTGCCAAGAAAGATG 1500

Qy  1519 CCAAGACTCCCTGCCACAGCAGCAGACCTGAAGCAGCGTCATTAGGTAATGGGGGAACAT 1578
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 CCAAGACTCCCTGCCACAGCAGCAGACCTGAAGCAGCGTCATTAGGTAATGGGGGAACAT 1560

Qy  1579 TA 1580
         ||
Db  1561 TA 1562
```

```
Qy    472 ATACAGATATAGAGGAATCGACACANTAATGCNATGCTTATGGAGGCAACAGCGCTGCTGTGCT 531

Qy    554 TCAATGCTGTTAACTGGATTGAGTGAGTCCAGCTCTTGGGAT-------              591
          ||||||||||||||||||||||||||||||||||||||||||
Db    532 TCAATGCTGTTAACTGGATTGAGTGAGTCCAGCTCTTGGGAT-------              591

Qy    592 --------                                                      591
Qy    592 --------                                                      591
Db    592 CAGGAGATATTGCTGTATATGCCACACAGGAAATGCTAGACCTACAGTGGAGTGGACGCAG  651

Qy    592 ------------------GGGCTTCGTGGGACAC                            607
Db    652 TAGCTCTGCTTAATTGGGCCAAATGCTTCCTTTAATTTGAACGAGGGCTTCGTGGGACAC  711

Qy    608 ATATGCAACACATGCCTATGACTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG  667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    712 ATATGCAACACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG  771

Qy    668 ATGGAAAAACTCTCCATACACAGTGCTACCTCAGTGCATTAGACCCGTGCTATTCTGTCTACT  727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    772 ATGGAAAAACTCTCCATACACAGTGCTACCTCAGTGCATTAGACCCGTGCTATTCTGTCTACT  831

Qy    728 GCAAAAAAGATCCATGCCCAGTGGGCAGAAGAGGGAAATGATAAAGATTTTACCTTGAANG  787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    832 GCAAAAAGATCCATGCCCAGTGGGCAGAAGAGGGAAATGATAAAGATTTTACCTTGAATG  891

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAGAAATCTCTAGCTC  847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    892 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAGAAATCTCTAGCTC  951

Qy    848 GGANGTTGCTTGATGACTTCCTTAATGACCAGATTAGAATTAGTGTATCTATAGTG  907
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    952 GGATGTTGCTTGATGACTTCCTTAATGACCAGATTAGAATTAGTGTATCTATAGTG  1011

Qy    908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGCACCTACTTTGGATAGAGATGGAGA  967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1012 GCCTGGAAGCCTTTGGGGATGTTAAGTTAGGAAGCACCTACTTTGATAGAGATGGAGA  1071

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGAAAAACATCTTTACTTG  1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1072 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGAAAAACAATCTTTACTTG  1131

Qy   1028 TATCAAAATCAAAAATGGGAAAATATGTACACANTCTTCAGTATTGTCCTTGCATCTGTTC  1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1132 TATCAAAATCAAAAATGGGAAAATATGTACACATCTTCAGTATTGTCCTTGCATCGTTC  1191

Qy   1088 TAGCACACGTACTCACCTCAGCAATTAGCAGGGAGGAAGAGAATTGGAGTGTTTTCTTATGGTT  1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1192 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAGGAAGAGAATTGGAGTGTTTTCTTATGGTT  1251

Qy   1148 CTGGTTTGGCTGCCACTCTCGTACTCTTCTTAAAGTCACCAAGATGCTACACCGGGGTCTG  1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1252 CTGGTTTGGCTGCCACTCTCGTACTCTTCTTAAAGTCACCAAGATGCTACACCGGGGTCTG  1311

Qy   1208 CTCTTGATAAAATAAACAGAGCAGTTTATGTTGTATCTTAAAATCACAAGCCTTGATTCAAGAACTG  1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1312 CTCTTGATAAAATAACAGCAGTTTATGTCGTATCTTAAAATCACAAGGCTTGATTCAAGAACTG  1371

Qy   1268 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG  1327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1372 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTCGG  1431

Qy   1328 TCAACTATATTCCCCAGGGGTTCATTCAAACATTGCGTCCCACTCCACATTGATGAAACTG  1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1432 TCAACTATATTCCCCAGGGGTTCATTCAAACATTGCGTCCCACTCCACATTGATGAAACTG  1491

Qy   1388 GGGTGGACTGAAAAAGCACAGAAGAACTTAGGCTCGGCGTCCCACTCCCCAAATGAGACACTT  1447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1492 GGGTGGACTGAAAAAGCACAGAAGAACTTAGGCTCGGCGTCCCACTCCCCAAATGAGACACTT  1551

Qy   1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATTAGCAACTGAGCCATATATTCCAAGCCCCTG  1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1552 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATTAGCAACTGAGCATATATTCCAAGCCCTG  1611

Qy   1508 CCAGCAAAAGTACCAAGACNTCCTCGCCACAGCAGCAGCGACGAACCTGAAGCGAGCGCTGTCATTAGTA  1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1612 CCAGCAAAAGTACCAAGACTCCCTGCTACACAGCGACGAACCTGAAGCCAGCAGTTATTAGTA  1671

Qy   1568 ATGGGGAACATTAA  1581
          |||||||
Db   1672 ATGGGGTATGGTAA  1685
```

RESULT 7
BC029693
LOCUS   BC029693   3260 bp   mRNA   linear   ROD 07-OCT-2003
DEFINITION   Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
             (cDNA clone MGC:36525 IMAGE:5375374), complete cds.
ACCESSION   BC029693
VERSION   BC029693.1 GI:20988708
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3260)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22288257
    PUBMED   12477932
REFERENCE   2 (bases 1 to 3260)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Baylor College of Medicine Human Genome
             Sequencing Center
             Center code: BCM-HGSC
             Web site: http://www.hgsc.bcm.tmc.edu/cdna/
             Contact: amg@bcm.tmc.edu
             Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleeged, H.,
             Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
             A.N., Gibbs, R.A.
             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 53 Row: p Column: 14
             This clone was selected for full length sequencing because it
             passed the following selection criteria: matched mRNA gi: 22123370.

```
FEATURES             Location/Qualifiers
source               1. .3260
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="MGC:36525 IMAGE:5375374"
                     /tissue_type="Mammary tumor, MMTV-LTR/INT3 model. 5 month
                     old mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
gene                 1. .3260
                     /gene="Hmgcs1"
                     /note="synonym: MGC36662"
                     /db_xref="LocusID:208715"
                     /db_xref="MGI:107592"
CDS                  79. .1641
                     /codon_start=1
                     /product="Hmgcs1 protein"
                     /protein_id="AAH29693.1"
                     /db_xref="GI:20988709"
                     /db_xref="LocusID:208715"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQDAELEKYDGVDA
                     GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSYDIGRLEVGTEIIIDKSK
                     SVKSNMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                     AIYANGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVRKKIRAQWQKEGKGDKDPTLNDPGFMIFHSPYCKLVQKSIA
                     RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELPNQKTKAS
                     LLVSNQNGNMYTSSVYGSLASVLAQYSPQLAGKRVGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITAASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                     ESESAVISNGEH"
misc_feature         115. .1485
                     /note="HMG_CoA_synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CD:pfam01154"
ORIGIN

Query Match          59.8%;  Score 1197.4;  DB 10;  Length 3260;
Best Local Similarity 78.7%;  Pred. No. 3.8e-250;
Matches 1582; Conservative  0; Mismatches 281; Indels 146; Gaps  7;

Qy      67 GGGGAGGCGCCGCCGGACTCGTCCTTCGTTGGGCGTCACTCCCTTTCCTCTGCTGCCGGGT 126
           |||||||| ||||||||||||||||||||||  |||||||||| ||||||  |||||
Db       1 GGGCGGGGAGGCAGCAGTCTCGGGGTCTCCCTTGCTTGGTCGTCGTTCTTCCAGGGTCTGATC 60

Qy     127 CACGCTTGCCTTTCACCATGCCTCGGATCACACTTCCTTCCTTTGAATGCAGAAGCTTGCTCGGCCA 186
           |||||||||||||||||||||  |||||||||||||||||| ||||||||||||| ||||||
Db      61 CCCTTTGGGCTCTTTCACCATGCCTCGGGTCACTCCTTCCTTGAATGCAGAGGCTTGCTCTGCCA 120

Qy     187 AAAGATGTGGGAATTGTTGCGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCA 246
           |||||||||||||||||||||  |||  | |||||||||||||||||  || |||||||
Db     121 AAAGAATGGGAAATCGGTTGCCCCTTTGAAATCTACTTTCCTTTCAATATGTCGATCAAGCT 180

Qy     247 GAGTTGGAAAAATGATGGTGTGAGATCTGGAAAAGTAACTACTTTCCTTCTCAATATGTTGGCGGCCC 306
           ||||||||||| |||||||| |||||||||| ||||  ||||| ||||||  |||| |||||
Db     181 GAGTTGGAAAATACAGATGGTGTTGTAGAATGCTGGAAAAGTATACCAATCGGCCTGGGGTTCAG 240

Qy     307 AAGAGTGGGCCTTCTGCACAGTAGAGAGAGATTATTAACTCTCTTCTTTGCATGACTGTGGGTTCAG 366
           |||| ||||||||||||||||  ||||||| |||| || |||||||||||| ||||||  |||||
Db     241 AGGATGGGGCTTCTGCACGGATCGTGAAGACAATCAACTCTCTTTTGCCCTGACTGTGGTTCAG 300

Qy     367 AATCTTATGGAGAGAGAAAAATAACCTTTCCTTTCCTTATAGATTGCAATTGGGGCGGCGGCTGGAAACA 426
           ||||| ||||||||||||||||| ||||||||||  ||||||| |||||||||||||||  ||||||
Db     301 AAACTGATGGAGAGACAGTAGCCTTTTCCTTATGATTCGATTGGGGGGGGCGCGGCTAGAGTTGGAACA 360

Qy     427 GAGACAATCATCGACAAATCGAAATCAGTGAAGTCTAATTGATGCAGCTGTTTGAACAG 486
           |||||||||| ||||||||||  |||||||| | ||||||||||||||||  |||||||
Db     361 GAGACAATCATCGACAAATCGAAATCAGTGAAGTCTAATTGATGCAGCTGTTTGAGGAG 420

Qy     487 TCTGGGGAATACAGATATTAGAACGACACAACTAAATGCATGCTATGGAGGCACAGCT 546
           ||||||||||||||||| ||||||||  |||||||||||||||||||||||||||| |||
Db     487 CACAGTTTGGATGAAGGAGTAGGACTTGTTGCATTCAAAACATAGCAACTGAGGCACATAATTCCA 1560
```

Qy   421 TCTGGGGAATACAGATATTAGAGAGGAATAGAATAGAATACAAAGCAATGACTATATGGGGGGGCACAGCT 480
Qy   547 GCTGTGCTTCTTCAATGCCTGTGTTAACTGGATTTGAGTTCCAGCTGTTCTTGGGGGAT----------- 591
Db   481 GCAGTCTTCAATGCCGTGAACTGGGTCGGAATCCAGCTCTTGGGCGATGGGATAGCTCTCTG 540

Qy   592 --- 591

Db   541 GTAGTTGCAGGAGGACATTGCTATATATGCCACAGGAAAATGCCAGAACCTTACAGGTGGAGTT 600

Db   601 GGAGCTGTGGCCCTCGCTAATTGGGCCAAAGCCTCCTCTTAAATTTTTGACCGGAGGGCTCCGT 660

```
Qy   1501 AGCCCTGCCAAGAAAGTACCAACAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTC 1560
          |||||||  ||||||||  |||  |||||||||||| |||   ||||  ||||  |||||||||
Db   1561 AGCCCTGTTAAGAAAGTGCCAAGACTCCCTGCAACCTCGGCCGAATCTGAAATCAGCTGTC 1620

Qy   1561 ATTAGTAATGGGGAACATTAAAGATACTCTGTGGAGGTGCAAGACTTCAGCGGTGGGGGGC 1620
Db   1621 ATCAGTAACGGGGAGGACACTCTGTGGAGTCTGGCCTTTACA----------GAGGCTCGGG 1669

Qy   1621 ATGGGGTGGGGGGTATGGGAACAGTTGGAGGAGGGGGATAATCTGGGGGATAATTTAAAGGA 1680
Db   1670 GCCGGGATGGGGTATGGGGAAACGGTTGGGAGGGAAT-GGAGTGTCTTGGGGACAATTTGCA-GA 1727

Qy   1681 TTACATGTTATGTATAAATTTTATGTGACTGACACATGGGAGTCTGGATGACTATATCGTGTACTT 1740
Db   1728 TTATGTGTTGCTTAAAATGTAATGTAACTGACACAGAGCCCAGAAAAGCTATTGTTGTTTTT 1787

Qy   1741 GGGAAAGTCT---CTTTGCTCATTTGCTGACATGCTTCCGTGTTCCTGTTGGTTGGTTGG... 1797
Db   1788 GGAAAAGTCTCTCGACTTGAACTTGCTAACAATTGCTAACATGGGAGTCTGGATGACTATATCCTTCCAATGC 1847

Qy   1798 CAAATGTACTCGAATGATGTT-AAGGGCCTCGTGTAAAACTTCATACCTCTTTGGGCCATTTG 1856
Db   1848 TAAAATGTACTGCAGTGAGTTGTTAAGGGCTGTTAAGGGCCTTTGTGAGACCTTATTCTGGGGTTTA 1907

Qy   1857 TATGCATAGTGATGTTGTTTGGTTTTTAAACATGGTATAATGAATTGTGTGTACTTCCTGTTCGTTGTTCAGAAGGA 1916
Db   1908 TACGGCATGCAGTTAGTTTTCAGAGTGTGGT---ATGAACTGAGTGCTTCTGACAGCAAAA 1964

Qy   1917 AGCAGGAGGTACTAATCCCAATTAAAAAA 1945
Db   1965 GGGAGGAGGTACTACTGTCCAATTTTAAA 1993
```

RESULT 8
BC034317

```
LOCUS       BC034317                3202 bp    mRNA    linear   ROD 12-NOV-2003
DEFINITION  Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36662 IMAGE:5366786), complete cds.
ACCESSION   BC034317
VERSION     BC034317.1  GI:21706865
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1  (bases 1 to 3202)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
   MEDLINE  22388257
REFERENCE   2  (bases 1 to 3202)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (02-JUL-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 53 Row: k Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 22122370.
FEATURES             Location/Qualifiers
     source          1..3202
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="MGC:36662 IMAGE:5366786"
                     /tissue_type="Eye, retina, mouse strain C57Bl/6"
                     /clone_lib="NIH_MGC_94"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..3202
                     /gene="Hmgcs1"
                     /note="synonym: MGC36662"
                     /db_xref="GI:21706866"
                     /db_xref="LocusID:208715"
     CDS             25..1587
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
                     1"
                     /protein_id="AAH34317.1"
                     /db_xref="GI:21706866"
                     /db_xref="LocusID:208715"
                     /translation="MPGSLPLNABACWPKDVGIVALEIYFPSQVDQAELEKYDGVDA
                     GKYTIGLGQARMGPCTDREDINSLCLTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                     AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDPYKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKILVQKSLA
                     RMFLNDFLNDQNDRDKNSIYSGLEAFGDVEKTYFDRVEKAPMKASSELFNQKTKAS
                     LIVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRVGVFSYGSGLAATIYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPFFNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                     ESSSAVISNGEH"
     misc_feature    61..1431
                     /note="HMG CoA synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CDD:pfam01154"
ORIGIN
```

```
Query Match             59.5%;  Score 1190.2;  DB 10;  Length 3202;
  Best Local Similarity   79.8%;  Pred. No. 1.4e-248;
  Matches 1555;  Conservative    0;  Mismatches  248;  Indels  146;  Gaps    7;

Qy    127 CACGCGTTGCTCTTTCACCCATGCCTGGATCACTTCCTTTGAATGCAGCAGAAGCTTGCTGGGCCA 186
          ||||| ||||||||||||| |||||||||| |||||||||||||||||||||| |||||||||| ||
Db      7 CCCTTTGGCTCCTTTCACCATGCCTGGTCACTTCCTTTGAATGCAGCAGAGGCTTGCTGGCCA 66

Qy    187 AAAGATGTGGGAATTGTGGCCCTTGACATCTATTTTCCTTCCTTTGAATGCAGAAGCTTGCTGGATCAAGCA 246
          ||||||||||||||| || ||||||||| |||| ||||||| ||  ||  |||||||  |||| |||||||| 
Db     67 AAAGATGTGGGAATCGTTGCCCTTGAAATCTACTTTCCTTCTCAAATATGTCGATCAAGCT 126
```

RESULT 9
LOCUS   BC023851   2703 bp   mRNA   linear   ROD 06-OCT-2003
DEFINITION Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36620 IMAGE:5347038), complete cds.
ACCESSION   BC023851
VERSION   BC023851.1 GI:2271493
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 2703)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

REFERENCE   2 (bases 1 to 2703)
  AUTHORS   Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2703)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (05-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K. Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L, Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 55 Row: h Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 22122370.
FEATURES   Location/Qualifiers
     source   1..2703
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:36620 IMAGE:5347038"
                     /tissue_type="Mammary tumor. C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene   1..2703
                     /gene="Hmgcs1"
                     /note="synonym: MGC36662"
                     /db_xref="LocusID:208715"
                     /db_xref="MGI:107592"
     CDS   149..1711
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
                     1"
                     /protein_id="AAH31851.1"
                     /db_xref="GI:23271494"
                     /db_xref="LocusID:208715"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA

ORIGIN
misc_feature   185..1555
                /note="HMG CoA synt; Region:
                Hydroxymethylglutaryl-coenzyme A synthase"
                /db_xref="CDD:pfam01154"

GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
SVKSNLMQLFEESGNTDIEGIDTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDPYKPDMLSEYPVVDG
KLSIQCYLSALDRCYSVRKKIRAQWQKEGKDDFTLNDFGFMIFHSPYCKLVQKSLA
RMFLNFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAPMKASSELFNQKTKAS
LLVSNQNGNMYTISSVYGGLIASVLAQYSPQQLAGKRVGVFSYGSGGLAAILYSLKVTODA
TPGSALDKITAISLDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
GTWYLVRVDEKHRRTYARPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
ESESAVISNGEH"

	Qy	134	GCTCTTTCACCCATGCCTGGATCACTTCCTTCTCAATATGTTGAATGCAAGAAGCTTGCTGGCCAAAAGATG 193
	Db	138	GCTCTTTCACCCATGCCTGGCTCACTTCCTTTGAATGCAAGAGAGCTTGCTGGCCAAAAGATG 197
	Qy	194	TGGGAATTGTTGCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCAGAGTGG 253
	Db	198	TGGGAATCGTTGCCCTTGAAATCTACTTCCTTCAATATGTCAGATAGTCGAGTTGG 257
	Qy	254	AAAAATATGATGGGTGTAGAGTGCTGAAAAGTATACCATTGGCTTGGGCCCCACAGATGG 313
	Db	258	AAAAAATACGATGGTGTAGAGTGCTGGAAAGTACCATCGGCCTGGGCCCCACAGATGG 317
	Qy	314	GCTTCTGCACAGATGAGAGGAGGAATATTAACTCTCTTCTTTGCATGACTGTGGGTTCAGAATCTTA 373
	Db	318	GCTTCTGCACGGATGCTGTGAAGACATCAACTCTCTTTGCCTGACTGTGGGTTCAGAAACTGA 377
	Qy	374	TGGAGGAGAAATAACCTTCCTTATGATTGCATGGGGGGGCCTGGGAGTTGGAACAGAGAGACA 433
	Db	378	TGGAGGAGACATAGCCTTCCTATGATTGCATTGGGGGGCCTAGAAGTTGGAACAGAGACA 437
	Qy	434	TCATCGACAAAATCAAAGCTCTTGTGAAGACTAATTTCAGTGCAGCTGTTGGAGTCTGGA 493
	Db	438	TCATCGACAAAATCGAAATCAGTGAAGCTGAATTGACGCAGCTGTTGGAGTCTGGA 497
	Qy	494	ATACAGATATAGAAGCGAATCGACACAACTAATGCATGCTATAGGAGGCCACAGCTGCTGTCT 553
	Db	498	ATACAGATATAGAAGAAATAGATACAACAATGCATGCTATGGGGGGCCCACCACTGCCAGTCT 557
	Qy	554	TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------- 591
	Db	558	TCAATGCCGTGAACCTGGGTCGAAATCCAGCTCTTGGGATGACGATATGCTCTGGTAGTTG 617
	Qy	592	--GGGCTTCGGTGGGACAC 607
	Db	618	CAGGAGACAATGCTATATATGCCACAGGAAAATGCGACCTACAGGTGGGACAGGCTG 677
	Qy	592	-- 591
	Db	678	TGGGCCCTGCTAAATGGGCCCAAACGCTCCTCTAATTTTGACCGAGGGACAC 737
	Qy	608	ATATGCAACATGCCTATGATTTTACAAGCCTGATATGCTATATATCCTATAGTAG 667
	Db	738	ACATGCAGCATGCCTATGACTTTAGCAAGCCTGACATGCTGTCGTCTCGGTCG 797
	Qy	618	CAGGAGACAATTGCTATATATGCCACAGGAAAATGCCAGACCTACAGGTGGGACAGGCTG 677
	Db	668	ATGGGAAAACTTCCCATACACTGTGCATTAGCACCGGCTGTTCTCTTGTCTTACT 727
	Qy	592	-- 591
	Db	798	ACGGAAAAGCTCTTCCAATACAGTGCTACCTCAGCCGCCTGGACACCTGTATTCTGTCTACC 857
	Qy	728	GCAAAAAGATCCATGCCCAGTGGCAGAAGAGGGGAAATGATAAAAGATTTTTACCTTGAATG 787
	Db	858	GCAAAAAGATCCGTGCCCACAGTGGGCAGAAAGAGGGAAAGAAATTTACCCTGAATG 917
	Qy	788	ATTTTGGCTTCATGATCTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 847

```
Db   918  ATTTTGCTTCATGATCTTTCACTCACCATATTGTGAACTGGTCGAGAAATCTCTAGCTC    977
Qy   848  GGATGTTGCTGAATGACTTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTG   907
Db   978  GGATGTTCCTGAATGACTTTCCTTAATGATCAAAAACAGAGTAAAAACAGTATTTACAGTG  1037
Qy   908  GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGGACCACCTACTTTGATAGAGATGTGGAGA  967
Db  1038  GACTGGAAGCCTTTGGGACGTTAAATTAGAAGGATACTTACTTTGACAGAGATGTAGAA   1097
Qy   968  AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTG  1027
Db  1098  AGGCATTTATGAAGGCTAGTTCTGAACTCTTCAACCAGAAAACAAAGGCTCTTTTACTTG   1157
Qy  1028  TATCAAATCAAAAATGGAAATATGTACACATCTTCAGTATATAGTGGTTCCCTTGCATCTGTTC  1087
Db  1158  TGTCTAATCAGAATGAAAATATGTACACATCTGTCTATTGGTTCCCTGCATCTGTTCC     1217
Qy  1088  TAGCCAGTACTCACCTCAGCAATTAGCAGGGAGAGAATTGGAGTGTGTTTTCTTATGGTT   1147
Db  1218  TGGCACAGTACTCACCTCACCTCAGCAGTTGGCAGGGAAGAGGGTTGGAGTGTGTCTTCTTACGGTT  1277
Qy  1148  CTGGTTTGGCTCGCCACTCTGTTACTCTTTAAAGTCACACAGATAGTGTACACCGGGTCTG  1207
Db  1278  CTGGCTTGGCTCGCCACGCTGTACTCCCTTAAAGTCACACAGATAGCGCCACCAGGATCTG  1337
Qy  1208  CTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG  1267
Db  1338  CCCTTGATAAAAATAACAGCAAGTTTATGTGACCTTAAAATCAAGGCTTGACTCGAGACTT  1397
Qy  1268  GTGTGGCCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCCATCGAGAGA  1327
Db  1398  GTGTGGCCACCGGATGTCTTTGCTGAAAACATGAAGCTCAGAGAGGACACACATCACTTAG  1457
Qy  1328  TCAACTATAATTCCCCAGGGGTTCAATAGATTCACTCTTTGAAGGAAGGTGGTACTTAGTTA  1387
Db  1458  CCAACTATATTCCCCAGTGTTCAATAGACTCACTCTTTGAAGGAACGTGGTAATCTCGGTCA  1517
Qy  1388  GGGTGGATGAAAAGCACACAGAAGAACTTACGCTCGGGTCCCACTCCAAATGATGACACTT  1447
Db  1518  GAGTGGATGAAAAAACACAGAAGGACTTACGCCCGGCCCCTTCACCAAATGACCACAGTT   1577
Qy  1448  TGGATGAAGGAGTAGGACTTGTGCATTCAAACATGAGCATATTCCAAGCCCTG           1507
Db  1578  TGGATGAAGGAGAATGGGGGCTTCGTGCTGCATAGTAACGACAGCGAATATTCCAAGCCCTG  1637
Qy  1508  CCAAGAAAGTACCAAGACTCCCTGCCCACAGCAGCGAACCTGAAGCCAGCTGTCATTAGTA  1567
Db  1638  CTAAGAAAAGTGCCAAGACTCCCTGCCCAACCTCGGCCGATACTCTGAATCAGTCATCAGTA  1697
Qy  1568  ATGGGGAACAATTAAGAGATACTTCGTTGTGAGGTTCAAGGACTTCAGGGTGGGGTGGGCATGGGGT  1627
Db  1698  ACGGGGAGCACTGAGCAGTCTGTGGCCGTTTACA--------GAGGCTCGGGGCGCCGGGA    1746
Qy  1628  GGGGGGTTATGGGAACAGTTGGGAGGAATGGGGATAATCTGGGGATTAAGGATTACATG    1687
Db  1747  TGGGGTTATGGAAACGGTTGGAGGAAT-GGATGTGTCTGGGACAATTTTGCA-GATTAGTG  1804
Qy  1688  TTATGTAAAATTTTTATGTGACTAGCAATGGTGTGACTATGTCTGTACTTGTTAGGGAAAG  1747
Db  1805  TTGCTTAAAATGTAATGTAACTGACACAGAGCCCAGAGCTAAGGCTAGTATCCTGTACTTCA  1864
Qy  1748  TCT---CTTTGCTCTATTTGCTGACATGCTTCCTGTTGTCCTGTGTCTGGCTCCCAATGCCCAAATGT  1804
Db  1805  TTGCTTAAAATGTAATGTAACTGACACAGAGCCCAGAGCTAAGGCTAGTATCCTGTACTTCA  1864
Qy  1865  TCTCTCGTGACTGAACTGCTAACACATGCTTCCTCGTGTTTGTGTCTGGCTCGGCCAGTGGTTACATGT  1924
Db  1925  ACTGCAGTGATGTTAAAAGGGCTCTGTAGAAACTTATACCTCTCGGATGTTTATACCGCAT  1984
Qy  1805  ACTCGAATGATGTT-AAGGGCTCTGTAAAAACTTCATACCTCTTGGCCATTGTATGCGAT    1863
Db  1925  ACTGCAGTGATGTTAAAAGGGCTCTGTAGAAACTTATACCTCTCGGATGTTTATACCGCAT  1984
Qy  1864  GATGTTTGGTTTTTAAACATGGTATAATGAATTGTGTTGTACTTCTGTCAGAAGAACAGAG  1923
Db  1985  GCAGTTTTAGTTTTCAGATGTGGT---ATGAACTGAGTGCTTCTGACAGCAGAAAAGGCAGAG  2041

Qy  1924  GTACTAATCTCCAATTAAAAAA  1945
Db  2042  GTACTAGTGTCCAATTTTTAA   2063
```

RESULT 10

BC031363

```
LOCUS       BC031363                3278 bp    mRNA    linear   ROD 07-OCT-2003
DEFINITION  Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36633 IMAGE:5357048), complete cds.
ACCESSION   BC031363
VERSION     BC031363.1  GI:21618632
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3278)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE  22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3278)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbc-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 54 Row: e Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 22122370.
FEATURES             Location/Qualifiers
     source          1..3278
                     /organism="Mus musculus"
                     /mol_type="mRNA"
```

```
                    /db_xref="taxon:10090"
                    /clone="MGC:36633 IMAGE:5357048"
                    /tissue_type="Eye, retina, mouse strain C57Bl\6"
                    /clone_lib="NIH_MGC_94"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-SPORT6"
     gene           1. .3278
                    /gene="Hmgcs1"
                    /note="synonym: MGC366662"
                    /db_xref="LocusID:208715"
                    /db_xref="MGI:107592"
     CDS            81. .1643
                    /codon_start=1
                    /protein_id="AAH31363.1"
                    /db_xref="GI:21618633"
                    /db_xref="LocusID:208715"
                    /translation="MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELKKYDGVDA
                    GKYTIGLGQARMGFCTDREDINSLCPTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
                    SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                    AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                    KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFPLNDFGPMIFHSPYCKLVQKSLA
                    RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFNQKTKAS
                    LLVSNQNGNMYTSSVSGSLASVLASVLAQYSGSGLAGKRVGVFSYGSGLAATLYSLKVTDDA
                    TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDLFE
                    GTWYLVRVDEKHRTVARPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                    ESESAVISNGEH"
     misc_feature   117. .1487
                    /note="HMG CoA synt; Region:
                    Hydroxymethylglutaryl-coenzyme A synthase"
                    /db_xref="CDD:pfam01154"
ORIGIN

Query Match    59.3%;   Score 1188;   DB 10;   Length 3278;
Best Local Similarity   79.9%;
Matches 1551;  Conservative  0;  Mismatches 245;  Indels 146;  Gaps 7;
```

(Nucleotide alignment block — Qy/Db paired rows, positions 134–592 / 70–609 and continuing 592–1630 / 607–1678)

```
RESULT 11
AX700129
LOCUS       AX700129          3275 bp    DNA     linear   PAT 03-APR-2003
DEFINITION  Sequence 15 from Patent EP1284298.
ACCESSION   AX700129
VERSION     AX700129.1  GI:29536069
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Brooksbank,R.A., Dixon,A.K., Lee,K. and Pinnock,R.D.
  TITLE     Identification and use of molecules implicated in pain
  JOURNAL   Patent: EP 1284298-A 15 19-FEB-2003;
            WARNER-LAMBRT COMPANY (US)
FEATURES             Location/Qualifiers
     source          1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
                     /note="Cytosolic 3-hydroxy 3-methylglutaryl coenzyme A
                     synthase"
ORIGIN

Query Match     59.0%;  Score 1180.6;  DB 6;  Length 3275;
Best Local Similarity   78.3%;  Pred. No. 1.7e-246;
Matches 1559;  Conservative   0;  Mismatches  284;  Indels 148;  Gaps  6;
```

(Sequence alignment of Query (Qy) against Database (Db), shown in paired rows with position numbers.)

```
RESULT 12
LOCUS       AX827303                3275 bp    DNA     linear   PAT 12-DEC-2003
DEFINITION  Sequence 37 from Patent EP1344834.
ACCESSION   AX827303
VERSION     AX827303.1  GI:39837392
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Boess,F.; Suter-Dick,L. and Wolf,D.
  TITLE     Methods for the toxicity prediction of a compound
  JOURNAL   Patent: EP 1344834-A 37 17-SEP-2003;
            F. HOFFMANN-LA ROCHE AG (CH)
FEATURES             Location/Qualifiers
     source          1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
ORIGIN

Query Match  59.0%;  Score 1180.6;  DB 6;  Length 3275;
Best Local Similarity  78.3%;
Matches 1559;  Conservative 0;  Mismatches 284;  Indels 148;  Gaps 6;
```

```
Qy   1388 GGGTGGGATGAAAAGCACAGAAGAACTTACGACTCGGCGTCCACTCCAAAATGAATGACACTT 1447
          |||||||||||||||||||||||||||||||||||| ||||| |||| |||||||| | ||||
Db   1461 GAGTGGGATGAAAAGCACAGAAGAACTTACGACCCGGCGTCCCTCCACAAAATGACCACAGTT 1520

Qy   1448 TCGGATGAAGGAGTAGGACTTGTGCCATTCAAACATAGCAACTGAGCAGCATAATTCCAAGCCCTG 1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1521 TCGGATGAAGGAGTGGGACTTGTGCCATTCAAACACACAGCTACAGACGCATCATCCAAGCCCCG 1580

Qy   1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCGACGAGAACCTGAAGCAGTGTCATTAGTA 1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1581 CTAAGAAAGTGCCAAGACTTCCTGCCAGAGCGAACCTGAGTCGTCGTTCATCAGTA 1640

Qy   1568 ATGGGGGACACATTAAGATACTCTGTTGAGGTTGCAAGACTTCAGGGTGGGGGTGGGGGT 1627
          ||||||||||||||||||||||||||||||||||||||||
Db   1641 ACGGGGGACCACTGGAGCCGTGGCCTTCACGAGGCTCGGGGGT------------GGA 1687

Qy   1628 GGGGGGTATGGGGAACAGTGGGGAATGGGGAATATCTGGGGGGGATAATTTAAAGGATTACATG 1687
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1688 TGGGGGGTATGGGGAAACGTTAGAGGAAT--GGATGTCTTGGGACAAATTTACAGGTTACGGTGT 1746

Qy   1688 TTATGTAAAATTTTTATGTGACTGACATGGAGCCTGGACTGAAGTCGTTGTACTTGGGAAAG 1747
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1747 TGCTTAAAAATGTAAATGTAACTGACACACAGAGCCCAGAAAAACGTTGTTGTGTTTTTGGAGAG 1806

Qy   1748 TCTCTTTGCTCTATTTGCTGACATGCTTCCTGCTGGTCGGCCAATGCCAAATGTTACT 1807
          |||||||||||||||||||||||||||||||||||||||||| ||| |||
Db   1807 TCTCGCTG----AGCTTCCTGCTGTGGGGCTGGCCAATGGTCAATGGTGAATGTACT 1862

Qy   1808 CGAAATGATGTTAAGGCTTCGTGAAAACTTCATCTTTGGCCATTTGTTATGTCATGGATG 1867
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 GGCATGGGTGTTAAGGGCTTCTGCAGAAGCTCATGGTCGGTTGGGATGACGAGAGGTAC 1922

Qy   1868 TTTGGGTTTTTAAACATGGTATAATGAATTGTGTACTTCTGTCGAAAGAAAAGCAGAGGTAC 1927
          |||||||||||||||||||||||||||||||||||||||||||
Db   1923 GTTAGGGCTTCAAAACTCGGT---CTGAACTGAGTCGCTTCTGACTGCA--AAGGCAGAGGTAC 1978

Qy   1928 TAATCTCCAATTAAAAAAATTTTTAAACATTGTAAAAAAAAAAAAAAAAAAAAAAAAAA 1987
          |||||| |||||||||||||||||||||||||||||||||||||||||
Db   1979 TGCTGTCCAGTTTAAAAAATGTTTTTTTTTTTTTTTAATGTGTAAGGAATTTTTTTATACTTA 2038

Qy   1988 AAAAAAAAAAA 1998
          ||||||||||
Db   2039 AATAAAAAAAA 2049
```

```
Qy    134 GCTCTTTCACCCATGCCTGGATCACTTCCTTCCTTTGGAATGCAGAAAGATG 193
          |||||||||||||||||||||||||| |||||||||||| | ||||||||||
Db     81 GCTCTTTTCACCCATGCCTGGGTCACTTCCTTCCTTTGAAATGCAGAGGTTGG 140

Qy    194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTTCAAATATGTTGATCAAGCAGAGTTGG 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 TGGGGAATCGTTGCCCTTGAAATCTACTTTCCTTCTTCCAGTATTGGATTGG 200

Qy    254 AAAAATATGATGGTGTGAGATGCTGGAAAGTATACCATTGGCTTGGGCCAAGATGG 313
          ||||||||||||||||||||||||||||||||||||||||||||
Db    201 GCTTCTGCACGGGATGCGCGAGACCGTGAGGTCTGGGGCCAGGAAACTGA 260

Qy    314 GCTCTTCTGCACAGAATAGAGAGAGATATTAACTTCCTTCTTGTGCTGTCTTCAGAATCTTA 373
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 TCATCGACAAATCAAAGTCTGTGAGACTAAATTTGATGGCGCGGTCTGGCCAGGAGCGA 320

Qy    374 TGGAGAGAAATAACCTTCCTATAGTTGCATTGGGGGGCTGGAGTTGGAACAGAGACAA 433
          |||||||||||||||||||||||||||||||||||||||||||
Db    321 TGGGAGAGAAATAGCCCTTCCTATGATCAGTTGGCCGGCTGGGGACGAGAAGGACAA 380

Qy    434 TCATCGACAAATCAAAGTCTGTGAGAGACTAAATTTGATGGCGCGGTCTGGCCAGGAGCTGGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 TCATCGACAAATCAAAGTCTGTGAGACTAAATTTGATGGCGCGGTCTGGCCAGGAGCTGGGA 440

Qy    494 ATACAGATATAGAGGAATGCATGCGGCACAGCTGTTGTCTGTCT 553
          |||||||||||||||||||||||||||||||||||||||
Db    441 ATACAGATATAGAGGAATGCATGCGGCACAGCTGTTGGGGGCAGCCGCAGTCT 500

Qy    554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGGAT------------------- 591
          |||||||||||||||||||||||||||||||||||
Db    501 TCAATGCTGTTGAACTGGATCCAGCTCTTGGGATGGACGATACGCTTTGGGTAGTTG 560

Qy    592 -------------------------------------- 591
          
Db    561 CAGGAGACATCGCTATATATGCCTCAGGAAACGCCAGGCCTACAGGTGGGAGCTG 620

Qy    592 ----------------------------------GGGCTTCGTGGGAGAC 607
                                            |||||||||||||||
Db    621 TGGCCTCTGCTAATTGGGCCAAATGCTCCTGTAAATTTTTGACCGAGGCTTCGTGGGAGAC 680

Qy    608 ATATGCAACATGCCTATGATTTTACAAGCCTGATATGCTATCTGAAATATCCTATAGTAG 667
          |||||||||||||||||||||||||||||||||||
Db    681 ACATCGCAGCATGCCTACGACTTTTACAAGCCTGACATGCTCTCTGAAATACCGTTGGTAG 740

Qy    668 ATGGAAAAACTCCCATACACATGCTGCTTACCTCAGTGCATTAGACCGCTGCTACTGTCTACT 727
          ||||||||||||||||||||||||||||||||||||||||
Db    741 ACATGCAGCATGCCTACGACTTTTACAAGCCTGACATGCTCTCTGAATACCCTGTGGTAG 800

Qy    728 GCAAAAAGATCCATGCCCAGTGGCGAGAAGAGGCGAAATGATAAAGATTTTTACCTTGAATG 787
          ||||||||||||||||||||||||||||||||||||||
Db    801 GCAAAAAGATCCGGGCCGCCCAGTGGCCAGAAAGAGGAGAAAGGATAAAGATTTTTACCCTGAATG 860

Qy    788 ATTTTTGGCTTCACTATTGGGACCATTCACCTCGCCATACTGTAACCTGGTCAGAAATCTCTAGCTC 847
          ||||||||||||||||||||||||||||||||||||
Db    861 ATTTTTGGCTTCATGGTTCTTTGCACGAAATGGTGTGAAACTGGTTGCAGAAAATTCTAGCTA 920

Qy    848 ATGGAAAAACTCCCATACACATGCTACCTGCTACTCAGTGCATTAGACCGCTGCTACTGTCTACT 907
          ||||||||||||||||||||||||||||||||||||||||
Db    921 GGATGTTCCTGAATGACTTTCTTACGGATCAAAAACAGAGTATTTTTTATTCTCAGTAGTG 980

Qy    908 GCCTGGAAGCCTTTGGGAGTTAAATTAGAAGAGACCACCTACTTTGATAGAGGCAGATGTCGGAGA 967
          |||||||||||||||||||||||||||||||||||||||||
Db    981 GGCTCGGAAGCCTTTGGGAGTGAAATTAGAAGAAGTACTTACTTTGAAA 1040

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGTCAGAAACAAAGGCATCTTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||
Db   1041 AGGCATTTATGAAAGGCTAGTGCTGACCTAATCAACGAGAAAACCAAAGGCATCTTTACTTG 1100

Qy   1028 TATCAAATCAAAATGTACACATCTTCAGTATATGGGTCCCCTTGCACTCGTTC 1087
          |||||||||||||||||||||||||||||||||||||||||||
Db   1101 TATCGAATCAAAATGGAAAACATGTACACATCCTCTGTATAGGGTTCCTTGCACATCGTTC 1160
```

```
RESULT 13
LOCUS       RNCHMCA              3275 bp    mRNA    linear   ROD 18-JUN-2003
DEFINITION  Rattus norvegicus mRNA for cytosolic 3-hydroxy 3-methylglutaryl
            coenzyme A synthase (EC 4.1.3.5).
ACCESSION   X52625
VERSION     X52625.1  GI:55946
KEYWORDS    cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Ayte,J., Gil-Gomez,G. and Hegardt,F.G.
  TITLE     Nucleotide sequence of a rat liver cDNA encoding the cytosolic
            3-hydroxy-3-methylglutaryl coenzyme A synthase
  JOURNAL   Nucleic Acids Res. 18 (12), 3642 (1990)
   MEDLINE   90301491
    PUBMED   1972979
  REMARK    (b 1-38,40,41-62,64-3275)
REFERENCE   2
  AUTHORS   Hegardt,F.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-APR-1990) Hegardt F.G., University of Barcelona, Unit
            of Biochemistry, School of Pharmacy, Placa Pius XII, s/n. 08028
            barcelona, Spain
  REMARK    (b 1-38,40,41-62,64-3275)
REFERENCE   3  (bases 1 to 3275)
  AUTHORS   Haegardt,F.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1990)
FEATURES             Location/Qualifiers
     source          1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /clone="lambda-CCS1"
                     /tissue_type="liver"
     old_sequence    31..70
                     /note="revised by [3]"
                     /citation=[3]
     CDS             92..1654
                     /citation=[3]
                     /EC_number="4.1.3.5"
                     /codon_start=1
                     /product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A
                     synthase"
                     /protein_id="CAA36852.1"
                     /db_xref="GI:55947"
                     /db_xref="GOA:P17425"
                     /db_xref="SWISS-PROT:P17425"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                     GKYTIGLGQARMGFCPTDREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTEPIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
                     AIYASGNARPTGVGAVALLIGPNAPVIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQNRDKQNSIYSGLEAGLDVKLEDTYFDRDVEKAFMKASAELFNQKTKAS
                     LLVSNQNGNMYTSSVYSSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWLLVRVDEKIRRTYARRPSTNDHSLDEGVGLVHSNTATEHIPSPAKKVPRLPATSG
                     EPESAVISNGEH"
ORIGIN

Query Match             59.0%;  Score 1180.6; DB 10; Length 3275;
Best Local Similarity   78.3%;
Matches 1559;  Conservative    0;  Mismatches  284;  Indels  148;  Gaps   6;
```

```
Qy      134 GCTCTTTCACCATGCCTGGATCACTTCACTTCCTTTGGAATGCAGAAGCTTGCTGGCCACATG 193
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       81 GCTCTTTCACCATGCCTGGGCTCACTTCCTTTGAAATGCAGAAGGCTTGCTGGCCAAAGATG 140

Qy      194 TGGGAAATTGTTGCCCCTTGACGATCTATTTTTCCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      141 TGGGAATCGTTGCCCCTTGAAATCTACTTTCCTTCTCAGTATGTTGATCAAGCAGAGTTGG 200

Qy      254 AAAAAATATGATGGGGTGTAGATGCCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGAGATGG 313
            |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db      201 AAAAATACGATGGTGTAGATGACTGGAAAAGTATACCATTGGCCTGGGCCAGGCCAAGATGG 260

Qy      314 GCTTCTGCACAGATAGAAGAAGAGATATTAACTCTCTTTGCATGACTGTGGTTCAGATCATTA 373
```

```
Db   261 GCTTCTGACGGATCGCGAAGACATCACTCTCTCTGCCTGACTGTGGTTCAGAAACTGA
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 TGGAGAGAAAATAGGCTTTCCTATGATTGATTGGGGCGGTGGAAGTTGGAACAGAGACAA 380
Qy   374 TGGAGAGAAAATAACCTTTCCTATGATTGATTGGGGCGGTGGAAGTTGGAACAGAGACAA 433
Db   381 TCATCGACAAAAATCAAAAATCCGTGAAGTCTTAATTTGATGATGCAGCGTCTTTGGGA 440
Qy   434 TCATCGACAAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGCGTCTTTGAGAGTTCTGGGA 493
Db   441 ATACAGAATATAGAAGGAAATACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCT 500
Qy   494 ATACAGATATAGAGAGGAAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCT 553
Db   501 TCAATGCTGTTGAACTGGATCGAATCCAGCTCTTGGGA 591
Qy   554 TCAATGCTGTTAACTGGATTGAGTTCCAGCTCTGGGAT 591
Db   561 CAGGAGACATCGCTATATATGCCTCAGGAAACGCCAGGGTGAGTTGGAGCTG 620
Qy   592 -------------- 591
Db   621 TGGCTCTGCTTAATTGGGCGAAAATGCTTCCTGTAATTTTGACCGAGGGCTTCGTGGGAC 680
Qy   592 --------- GGGCTTCGTGGGGACAC 607
Db   681 ACATGCAGCATGCCTTACGCAGACTTTTACAAGCCTCTCTGAATAACCTGGTAG 740
Qy   608 ATATGCAACATGCCTCTATGTCTCTATCTTACAGAAGATTCTATATAGTAG 667
Db   741 ATGGAAAACTCTCCATACAGTGCTATACCGCTGCCCTGCTATTCGTGTCTACC 800
Qy   668 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCCAATAGAGACCGCTGCTACTCTGTCTCACT 727
Db   801 GCAAAAAGATCATCCCGGCCCAGTGGCAGAAAATGGGAAATAAAGATTTTACCCTGAATG 860
Qy   728 GCAAAAAGATCCATGCCAGTGGGCAGAAAATGAGGGAAATGATAAAGATTTTACCTTGAATG 787
Db   861 ATTTTGGCTTCATGATTCTTCACTCACCATATTGTGGTTCAGAAAATCCTAGCTA 920
Qy   788 ATTTTGGCTTCATGATTCTTTCACTCAGTCTTATGTAAACTGGTTTCCTTATAGTAG 847
Db   921 GGATGTTCCTGAATGACTTTCTTAACGATCAAAACAGTATTTACAGTG 980
Qy   848 GGATGTTGCTGATGACTTCCTTAATGACCAGAATAGAGGATAAAAATAGTATCTATAGTG 907
Db   981 GGCTGGAAGCCTTTGGGGAGTGAAATTAGAGAAGATACTTACTTCGAGGATGTGGAAA 1040
Qy   908 GCCTGGAAGCCTTTGGGAGTGTTAAATTAGAAGAGCACCTACTTTGATAGAGTGTGGAGA 967
Db  1041 AGGCATTTATGAAGGCTAGTGCTGAGCTATTCCAGAAAACAAAGGCATCTTTTGCTTG 1100
Qy   968 AGGCATTTATGAAGGCTAGCTCTGAACCTCTTCAGTCAGAAAAAACAAAAGGCATCTTTACTTG 1027
Db  1101 TATCGAATCAAAAATGGAAAACATGTACACATCCCTGTATACGGTTCCCTGTTCTCGTTTC 1160
Qy  1028 TATCAAAATCAAAATGGAAAATATGTACACATCTTCAGTGTATATATGGTTCCCTTGCATCTCTGTTC 1087
Db  1221 TCAATGCTTGGCTGCCACACTCTACTCCCTTAAAGTCACACAAGATGCCACACCAGGATCTG 1280
Qy  1148 CTGGTTTGGCTTGCCACCTCTGTCTCTTAAAGTCACACAGAATTGGAGTGTTCTCTTACGGTT 1207
Db  1161 TGGCACAGTACTCACCTCAACCTGAACGGCGGAAGAGGAATTGGAGAGTGTTTTTCTTATGGTT 1220
Qy  1088 TAGCACAGTACTCACCTCAGCAATTAGACAGGGAGGAATTGGAGAGTGTTTTTCTTATGGTT 1147
Db  1281 CTCTTGACAAAATAACACGAAGTTTATGTGACCTTAAGTCAAGGCTTGACTCAGGAAGAGT 1340
Qy  1208 CTCTTGATAAAAATAACAGCAAGTTATGTGATCTGATTCAAGAACTG 1267
Db  1341 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTTCAGAGAGGACACACAATCACTTAG 1400
Qy  1268 GTGTGGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACAATCACTTAG 1327
Db  1401 CCAACTATTATTCCCCAGTGTTCAAATAGATTCACTCTTCTTGAAGGGAACATGGTATCTAGTCA 1460
Db  1328 TCAACTATATTCCCCAGGGGTTCAAATAGATTCACTCTTCTTGAAGGAACGTGGTTACTTAGTTA 1387
Db  1461 GAGTGGATGAAAAGCACAGAGAAGAACTTACGCGGTCCCCACAAAATGACCACAGTT 1520
Db  1388 GGGTGGATGAAAAGCACAGAGAAGAACTTACGCGGTCCCCACACTCCAAAATGATGACACTT 1447
Qy  1521 TGGATGAAGAGTGGGACTTGTGCATTCAAAACACACAGCAGCATAATTCCAAGCCGG 1580
Qy  1448 TGGATGAAGAGTAGGACTTGTGCATTCAAATCAAACATAGCAACTAGAACATATTCCAAGCCCTG 1507
Qy  1581 CTAAGAAAAAGTAGCCAAGACTTCCTTGGGGCTGTTCAATAGACTTCAGGATGTTCATCAGTA 1640
Qy  1508 TGGGTATTGGGGAAAAGTTGCCAGAACTTGCTTTAGCTGTTCAATAGATTCAGGATGTCATCAGTTA 1567
Db  1688 GGGGGTATGGGGAACAGTTGGAGGGAATGGGAATATCTGGGGGGATTAAATTTTAAAGGATTACATG 1747
Qy  1628 GGGGGTATGGGGAACAGTTGGAGGGAATGGGAATATCTGGGGGATTAATTTTTAAAGGATTACATG 1687
Db  1688 TGGGGTATGGGGAAACGTTAGAGGAAT-GGAGTGTTCGGAGACAATTTCAGGATTACGTGT 1746
Qy  1641 ACGGGGAGCACTGAGAGCCGTGGCCTTCACAGAGGCTCGGGGGGCT-------GGA 1687
Db  1807 TCTCGGCTG----AGCTCCTAACACACTTCCTGCTGTGGGGCTGGCAAATGGGAATGTACT 1862
Qy  1748 TCTCTTTGCTCTATTTGCTGACATCGTTCTCTGTTGTGGGGCTCTGGGCCAAATGTAGCTAT 1807
Db  1923 GTTAGGCTTCAAAACTCGGT---CTGAACTGAGTGCTTCTGACTGCA-AGGCAGAGGGTAC 1978
Qy  1868 TTTGGTTTTAAACATGGTAATAGGTATATACTTCCTGTGATTAATTTGCTTG 1927
Db  1863 GCGATGGTGTTAAGGCGTCTCGCAGAAGCGTCATACCTCGCTGCGCATGGCGG 1922
Db  1979 TGCTGTCCCAGTTAAAAAATTGTTTTTTTTTAAATGTGTAAGAAATTTTATACTTA 2038
Qy  1808 CGAATGATGTTAAGGGCTCTGTAAAAACTTCATACCTCCTTTGGGCATTTGTTGTATGCATGGATG 1867
Db  1927 TTTGGTTTTAAACATGGTAATAGGTATATACTTCCTGTGATTAATTTGCTTG 1927
Db  1988 AAAAAAAAAAA 1998
Db  2039 AATAAAAAAA 2049
```

variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-35L17 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES   Location/Qualifiers
source   1..127832
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="1"
      /clone="RP11-35L17"
      /clone_lib="RPCI-11.1"

ORIGIN

Query Match   55.1%;   Score 1103.2;   DB 9;   Length 127832;
Best Local Similarity   79.4%;
Matches 1465;   Conservative   0;   Pred. No. 1.2e-229;   Mismatches 198;   Indels 183;   Gaps   6;

Qy   245 CAGAGTTGGAAAAATATGAGTGGGTGTAGATGCTGGAAAGTAGTATACCATTGCCTTGGGCCAGG 304
Db   80799 CAAAGTTGGAAAAATATGTTGATATAGAGTGTGGAAAAGTATACCATTGCCTTGGGCCAGG 80858

Qy   305 CCAAGAATGGGCCTTCTGCACAGATAGAGAAGAAGATATTAACTCTCTTCTTTGCATGACTGTTC 364
Db   80859 CCAAGAATGGGCGTCCTGCACACATCGGAGAAGAGATCAACTCTCTTCTTTGCATGACTGTTC 80918

Qy   365 AGAATCTTATTGGAGGAGAAATAACCTTTCCTATGATTGCATTGGGCGGGCTGGGAAGTTGGAA 424
Db   80919 AGAATCTTATTGGAGGAGAAAAATCTTTCCCTACGGATTGCACTGGGCCGACTGGGAAGTTGGAA 80978

Qy   425 CAGAGACAATCATCGGACAAATCAAAGTCTGTGAGAGACTAAATTGATGCAGCTGTTTGAAG 484
Db   80979 CAGAGACAATCAAGGACAAATCAAAGTCAGTGAAGACTAAATTGATCGACTGTTTGAGG 81038

Qy   485 AGTCTCGGGAATACAGATATAGAAGGAATCGACACAACTAATGCATATGGAGGACACAG 544
Db   81039 AGTCTCGGGGATATAGAATATAGAAGGAACCGACACACTCAATGCTATGGAGCCCCAG 81098

Qy   545 CTGCTCGTCTTCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------ 591
Db   81099 CTGCTCGTCTTCAGTCGTGTTAACTGGATTGAGTCCAGCTCTTGGGATGACACAGTTTGCCC 81158

Qy   592 --------------------------------- 591
Db   81159 TGGAGTTTGTGAGGAGATATTGCTGTATATCCCACAGAGAAAAGTTAGACTTACACGGTTAGAG 81218

Qy   592 -------------------------GGGCTTC 598
Db   81219 TCAGAGCAGTAGCTCTGCTAATTCAGCCACTTTAATTTTTTGAACGAGGGCTTC 81278

Qy   599 GTGGGACACATATGCAACACGTGCCTATGCCTATTGATGTTTTATCCTGAATATC 658
Db   81279 GTGGGACACATATGCAACACGTGCCTATGCCTTTATGAGCTGATATGCTCTCTGAAATGTC 81338

Qy   659 CTATAGTAGATGGGAAAACTCTCCATACAGATGCTGCTACCTCAGTGCGTGCTGCTACT 718
Db   81339 CTATAGTAGGAGTAGGGAAAAGCCCTCCACACAGAGTGCTGCGCTTCAGTGCCGTGCCATTTA 81398

Qy   719 CTGTGTTCTACTGCAAAAGAAGATCCGTGCCCCAGTGGTGGCAGAAGAGGGAGAAATGATAAAGATTTTA 778
Db   81399 CTGTCGACTCGCAGAAAACGATCCGTGCCCCAGTGGCAGAAGAAGAGGGGAGAAATGATAAAGATTTTA 81458

Qy   779 CCTTGAATGAGTTTGGCTTCATGATCTTTCACTCACCATTATTGTAAACTGGTTCAGAAAT 838
Db   81459 CCTTGAATAAATTTGGGCTTCCTGGAATGTTTCACTCACCCATATTGTAAACTGGTTCAGAAAT 81518

Qy   839 CTCTAGCTCCGGATGTTTGCTGAATGACTTCCTTAAATGACCAGAATAGAGATAAAAAATAGTA 898
Db   81519 CTCTAGCTCCGGAATGTTGCGGAATGACTTCCTTAAATGACCACCATATTGACGAGATAAAAAATAGTG 81578

Qy   899 TCTATAGTTGCCTTGGAAGCCTTTGGGAGTGTTAAAATTAGGAAGACACCTACTTTGATAGAG 958
Db   81579 TTTAGAAATGTCCTGGAAGCCTTTGGGAATGTTAAAAATTAGGAACATCTATTTTGATACAG 81638

Qy   959 ATGTGGAGAAGGGCATTTTATGGAGGCTAGCCTGTGAACTCTTGTTTCAGTGCTAGAAAAACAAAAGGCAT 1018
Db   81639 ATGTGGAGGATGGCATTTAATGGAAGGGAATAGTTCCGAACTCTCCCCATCACAGAAAAACAAAAGGCAT 81698

Qy   1019 CTTTACTTTGTATCAAATCAAAATGGGAAATATGTCACACATCTTCAGTTATGGTTCCCTTG 1078
Db   81699 CTTTGC-TGTATCCAGTGAAATAACAGCACCCTGTACTCTCTTAAAGTTCCACCAGAATGCCACAC 81756

Qy   1079 CATCTGTTCTTAGCACAGACTACTCACCTCAGCAATTAGCAGGGAAAGAAGCTAGAGAGAGACACACC 1138
Db   81757 CAACTGTCCCAGGCCCAGTAGTCACCTCAGCAATGATCAGGGAAGAG-ACTGGAGTGGTTCT 81815

Qy   1139 CTTATAGGTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACAC 1198
Db   81816 CGTACCGGTTCTTCGGGTTCAGCGGTGCCACCTGTACTCTTAAAGTTCCACCAGATGCCACAC 81875

Qy   1199 CGGGGTCTGCTCTCTTTGATCAAAAATAAACAGCACCCTGTACTCTTGATTCAGTTAAATCAAAGGCTTGATT 1258
Db   81876 CATGGGTCTGCTGCTTGATAAAATAACAGCAAGTTTATGTTATTGGTTGATCTTAAATCAAAGGCTTAGCT 81935

Qy   1259 CAAGAACTGGTGGTGCCAACGATGTCTTCGCTGAAAATACAGCAAGCTTATGGGGACTGGTGTTC 1318
Db   81936 CAAGAACTTGTGTGGTGCCAGATGTCTTTGTTGAAAAACATGAAGCTCAGAGGAGGACACCC 81995

Qy   1319 ATCATTTGGTCAAACTATATTATTCCCCAGGGTTCAATAAGATTCACTCTTTTGAAGGGACGTGGT 1378
Db   81996 ATCACTTGGGTCAGAACTTCGTTCCCCGGGGGCCCAAGAGAATTCACTCTTGAAGAACATGGT 82055

Qy   1379 ACTTAGTTAGGGTGGATGAAAAAGCACACAGAAGAACTTACGGCTGCTCGGGGTCCCACTCCAAATG 1438
Db   82056 ACTTAGCTATGGTATGGGTGAAAAGGCACAGAAGAAGTTACACTC-TTCTCCAAATA 82115

Qy   1439 ATGACACTTGTGATAAGAGGTAGGACTTGTGTCATTCAAAACATAGCAAACTGCAGAATATTC 1498
Db   82116 ATGACACTTGGGAGTAGGAGTCAGGACTTCGGCATTCAAAACATAGTAACTGCAACATATTC 82175

Qy   1499 CAAGCCCTGCCAAGAAAATACCAAGAACTCCCTCCCACACAGCAGCAACCTGAAGCAGCTG 1558
Db   82176 CAAGCCCTGCTTAAGGACAGTGCCAAGAACTCCCACGCCAACAGCAAAACGGAGCAGCTG 82235

Qy   1559 TCATTTAGTAAATGGGGAAACATTAAGATAATATTCGTTGGAGGTGCAAGACTTCAGGGTGGGGGTG 1618
Db   82236 TCATTAGTTAAATGGGGAAACAATTAAGATATCTGGGGATAATTTTTAAAG 82295

Qy   1619 GCATGGGGTGGGGTATGGGAACAGTTGGAGAAATGGGGATATCTGGGGATAATTTTTAAAG 1678
Db   82296 G--GGATGG 82302

Qy   1679 GATTACACATGTTATGTAAAATTTTTATGTGACTGACATGGAGCCTGGAACATTATATCGTTGTAC 1738
Db   82303 GATTACACATTGTTGCTCAGGATTTTTTATGTGTGAGTGGACAATGGAACGGCCATCGTGTAC 82362

Qy   1739 TTGGGAAAGTCTCT-TTGCTTCCTATTTGCTTGACACATGCTTCCGTTGTGTGGCTGGCAAATGC 1797

```
Db    82363  TTGGGAAAGTCNCTTTTTCCTCAATTGCTGACATGTTTCCCACTGTGGTTCTGGCCAATGC
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     1798  CAAATGTACTCGAATGAATGATGTTAAGGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTTGT  1857
              ||||||||||||||||||||||||||||||||||||
Db    82423  CAAGTGCCCTTGAGTGATGTTAAGGGCTTAAGGGCTCTATAAAACTTCAGACCTCTTCCTAGCCATTCGT  82482
              ||||||||||||||||||||||||||||||||||||||||||
Qy     1858  ATGCATGATGTTTGGTTTTTAAACATGTATATGAATTGTGTACTTCGTTGTCAGAAGAA  1917
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    82483  ATACATGAAGTTTAGTTTTCACCATGTATATAAATGATGAATCCTGTCCTTCTGTTGTCAGAGGAAA  82542
              ||||||||||||||||||||||||||||||||||||
Qy     1918  GCAGAGGTACTAATCTCCAATTAAAAAAATTTTTTAACATGTAAAAA  1963
              ||||||||||||||||||||||||||
Db    82543  GCAGAGGTACTAATCTTCAATTAAATTTTTTTTTAAAACAGATAAGAA  82588
```

```
RESULT 15
LOCUS       AC021792/c              170834 bp    DNA     linear   HTG 04-MAY-2001
DEFINITION  Homo sapiens clone RP11-661L5, WORKING DRAFT SEQUENCE, 29 unordered
            pieces.
ACCESSION   AC021792
VERSION     AC021792.3  GI:8072559
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170834)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-661L5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170834)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,P., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Piere,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassilliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7463245.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
            Center project name: L4319
            Center clone name: 661_L_5
            -------------- Summary Statistics
            Sequencing vector: M13; M77815; 99% of reads
            Sequencing vector: Plasmid; n/a; %-0.f%% of reads
            0.5244755244755245Chemistry: Dye-terminator Big Dye; 100% of
            reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 149820 bases at least Q40
            Consensus quality: 160100 bases at least Q30
```

```
            Consensus quality: 164767 bases at least Q20
            Insert size: 174000; agarose-fp
            Insert size: 168034; sum-of-contigs
            Quality coverage: 3.2 in Q20 bases; agarose-fp
            Quality co.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*       1  1607: contig of 1607 bp in length
*    1608  1707: gap of 100 bp in length
*    1708  2937: contig of 1230 bp in length
*    2938  3037: gap of 100 bp in length
*    3038  4551: contig of 1514 bp in length
*    4552  4651: gap of 100 bp in length
*    4652  5804: contig of 1153 bp in length
*    5805  5904: gap of 100 bp in length
*    5905  7016: contig of 1112 bp in length
*    7017  7116: gap of 100 bp in length
*    7117  8323: contig of 1207 bp in length
*    8324  8423: gap of 100 bp in length
*    8424 10330: contig of 1907 bp in length
*   10331 10430: gap of 100 bp in length
*   10431 12576: contig of 2146 bp in length
*   12577 12676: gap of 100 bp in length
*   12677 15039: contig of 2363 bp in length
*   15040 15139: gap of 100 bp in length
*   15140 17696: contig of 2557 bp in length
*   17697 17796: gap of 100 bp in length
*   17797 20417: contig of 2621 bp in length
*   20418 20517: gap of 100 bp in length
*   20518 23021: contig of 2504 bp in length
*   23022 23121: gap of 100 bp in length
*   23122 25657: contig of 2536 bp in length
*   25658 25757: gap of 100 bp in length
*   25758 29388: contig of 3631 bp in length
*   29389 29488: gap of 100 bp in length
*   29489 34207: contig of 4719 bp in length
*   34208 34307: gap of 100 bp in length
*   34308 38295: contig of 3988 bp in length
*   38296 38395: gap of 100 bp in length
*   38396 42889: contig of 4494 bp in length
*   42890 42989: gap of 100 bp in length
*   42990 47895: contig of 4906 bp in length
*   47896 47995: gap of 100 bp in length
*   47996 54590: contig of 6595 bp in length
*   54591 54690: gap of 100 bp in length
*   54691 59447: contig of 4757 bp in length
*   59448 59547: gap of 100 bp in length
*   59548 65836: contig of 6289 bp in length
*   65837 65936: gap of 100 bp in length
*   65937 72591: contig of 6655 bp in length
*   72592 72691: gap of 100 bp in length
*   72692 83266: contig of 10575 bp in length
*   83267 83366: gap of 100 bp in length
*   83367 94803: contig of 11437 bp in length
*   94804 94903: gap of 100 bp in length
*   94904 110268: contig of 15365 bp in length
*  110269 110368: gap of 100 bp in length
*  110369 124647: contig of 14279 bp in length
*  124648 124747: gap of 100 bp in length
*  124748 140934: contig of 16187 bp in length
*  140935 141034: gap of 100 bp in length
*  141035 153563: contig of 12529 bp in length
*  153564 153663: gap of 100 bp in length
*  153664 170834: contig of 17171 bp in length.
FEATURES             Location/Qualifiers
     source          1..170834
                     /organism="Homo sapiens"
```

```
          /mol_type="genomic DNA"
          /db_xref="taxon:9606"
          /clone="RP11-661L5"
          /clone_lib="RPCI-11 Human Male BAC"
misc_feature   1..1607
          /note="assembly_fragment"
misc_feature   1708..2937
          /note="assembly_fragment"
misc_feature   3038..4551
          /note="assembly_fragment"
misc_feature   4652..5804
          /note="assembly_fragment"
misc_feature   5905..7016
          /note="assembly_fragment"
misc_feature   7117..8323
          /note="assembly_fragment"
misc_feature   8424..10330
          /note="assembly_fragment"
misc_feature   10431..12576
          /note="assembly_fragment"
misc_feature   12677..15039
          /note="assembly_fragment"
misc_feature   15140..17696
          /note="assembly_fragment"
misc_feature   17797..20417
          /note="assembly_fragment"
misc_feature   20518..23021
          /note="assembly_fragment"
misc_feature   23122..25557
          /note="assembly_fragment"
misc_feature   25758..29388
          /note="assembly_fragment"
misc_feature   29489..34207
          /note="assembly_fragment"
misc_feature   34308..38295
          /note="assembly_fragment"
misc_feature   38396..42889
          /note="assembly_fragment"
misc_feature   42990..47895
          /note="assembly_fragment"
misc_feature   47996..54590
          /note="assembly_fragment"
misc_feature   54691..59447
          /note="assembly_fragment"
misc_feature   59548..65835
          /note="assembly_fragment"
          vector_side:left
          clone_end:SP6
misc_feature   65937..72591
          /note="assembly_fragment"
          vector_side:left
          clone_end:T7
misc_feature   72692..83266
          /note="assembly_fragment"
misc_feature   83367..94803
          /note="assembly_fragment"
misc_feature   94904..110268
          /note="assembly_fragment"
misc_feature   110369..124647
          /note="assembly_fragment"
misc_feature   124748..140934
          /note="assembly_fragment"
misc_feature   141035..153563
          /note="assembly_fragment"
misc_feature   153664..170834
          /note="assembly_fragment"

ORIGIN

Query Match    55.1%;   Score 1103.2;  DB 2;   Length 170834;
Best Local Similarity   79.4%;   Pred. No. 1.2e-229;
Matches 1465;  Conservative   0;  Mismatches 198;  Indels 183;  Gaps  6;
```

Alignment (Db = database strand, Qy = query strand) with flanking positions:

Strand	Start	End
Db	120442	120383
Qy	305	364
Db	120382	120323
Qy	365	424
Db	120322	120263
Qy	425	484
Db	120262	120203
Qy	485	544
Db	120202	120143
Qy	545	591
Db	120142	120083
Qy	592	591
Db	120082	120023
Qy	592	591
Db	120022	119963
Qy	599	658
Db	119962	119903
Qy	659	718
Db	119902	119843
Qy	719	778
Db	119842	119783
Qy	779	838
Db	119782	119723
Qy	839	898
Db	119722	119663
Qy	899	958
Db	119662	119603
Qy	959	1018
Db	119602	119543
Qy	1019	1078
Db	119542	119485
Qy	1079	1138
Db	119484	119426
Qy	1139	1198
Db	119425	119366
Qy	1199	1258

```
Qy    245 CAGAGTTGGAAAATATGATGGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCCAGG 304
```

```
Db  119365  CATGGTCTGCTCGTTTGATAAAATAAACAGCAAGCTTATGGGATCTTAAATCAAGGTGTAGCTT  119306
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    1259  CAAGAACTGGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGGACACCC     1318
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119305  CAAGAACTTGTGTGGTACCAGATGTCTTTGTTTGAAAAACATGAAAGCTCAGAGAGGGACACCC  119246

Qy    1319  ATCATTTGGTCAAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAACGTGGT     1378
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119245  ATCACTTGGTCAACTCTCGTTCCCCGGGGCCCAAGAGATTCACTCTTTGAAGAAAACATGGT  119186

Qy    1379  ACTTAGTTAGGGGTGGGATGAAAAAGCCACAGAAGAACTTACGGCTCCGGCCGTCCCACTCCCAAATG  1438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119185  ACTTAGCTATGGGTGGGGTGAAAAAGCCACAGAAGAAGTTACACTCCGTGTCCTTCTCCAAATA  119126

Qy    1439  ATGACACTTTGGATGAAGGAGTAGGACTTGTGCAATTCAAACATAGCAACTGAGCATATTC     1498
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119125  ATGACACTTGGGATGAAGGAGTCAGACTTGTTGCATTCAAACACAGTAACTGATCATATTC  119066

Qy    1499  CAAGCCCTGCCAAGAAAAGTTACAAAGACTCCCCTCCCACGCAGAACCTGAAGCAGCTG     1558
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119065  CAAGCCCCTGCTAAGAACAGTGCCAAGACTCCCCTGCCACAGCAGCAGAAACGGGAGCAGCTG  119006

Qy    1559  TCATTAGTTAAATGGGGAACATTAAGATACTTCTGTGAGGGTGCAAGACTTCAGGGGTGGGGTGG   1618
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  119005  TCATTAGTTAATGGGGACACATTAAGATATTCTGTGAGGGTGCAAGGGCTTTGGTGTTGTGAGGGGTG  118946

Qy    1619  GCATGGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGATAATTTTAAG     1678
            ||                                                       |||
Db  118945  G---------------------------------------------------GGATGG  118939

Qy    1679  GATTACACATGTTATGTAAATTTTTATGTGACTGACATGAGCCCTGGATGACTATCGTGTTAC     1738
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118938  GATTACACATGTTGCTCAGATTTTTATGTGAGTGACATGGAGCCCTGGATGACCATCGTGTAC  118879

Qy    1739  TTGGGAAAGTCTCT-TTGCTCTATTTGCTCGACATGCTTCCGTGTTGGTGGTCTGGGCCAATGC     1797
            |||||||||||||| ||||||||||||||||||||||||||||||
Db  118878  TTGGGAAAGTCTCTTTTCCTCAATTTGCTGACARTGTTCCCACTGTGGTCTCTGGCCAATGC  118819

Qy    1798  CAAAAGTACTCGAATGATGTTAAGGGCTCGTGTAAAAACTTCATACCTCTTTGGCCATTTGT     1857
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db  118818  CAAGTGCGCTTGAGTGGATTGTTAAGGGCTTCTATAAAAACTTCAGACCTCTTCTAGCCATTCGT  118759

Qy    1858  ATGCATGATGTTTGGTTTTTTTAAACATGTGTATATGTAATTGTGTTACTTCTGTCAGAAGAAA     1917
            ||||||||||||||||||||||||||||||||||||||||||||||
Db  118758  ATACATGAAGTTTAGTTTTTCAACCATGGTATAATGAATCCGTCCTTCTGTTCAGGAGGAAA  118699

Qy    1918  GCAGAGGGTACTAATCTCCAATTAAAAAATTTTTAACATGTAAAAA     1963
            ||||||||||||||||||||||||||||||||||||   ||||
Db  118698  GCAGAGGGTACTAATCTTCAATTAAATTTTTTTTAAAACAGATAAGAA  118653
```

```
OM protein - protein search, using sw model

                    GenCore version 5.1.6
```

```
Run on:          June 24, 2004, 12:56:26 ; Search time 44 Seconds
                 (without alignments)
                 565.671 Million cell updates/sec

Title:
Perfect score:   2511
Sequence:        US-10-622-516-2
                 1 MPGSLPLNAEACWPKDVGIV.........PRLPATAAEPEAAVISNGEH 478

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters:   141681

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	2480	98.8	520	1	HMCS_HUMAN	Q01581 homo sapiens
2	2362	94.1	520	1	HMCS_RAT	P17425 rattus norv
3	2352	93.7	520	1	HMCS_CRIGR	P13704 cricetulus
4	2097	83.5	522	1	HMCS_CHICK	P23228 gallus gall
5	1513	60.3	508	1	HMCM_HUMAN	P54868 homo sapien
6	1494	59.5	508	1	HMCM_RAT	P22791 rattus norv
7	1493	59.5	480	1	HMCM_MOUSE	P54869 mus musculi
8	1466	58.4	508	1	HMCM_PIG	O02734 sus scrofa
9	1352.5	53.9	453	1	HMC1_BLAGE	blattella g
10	1272.5	50.7	455	1	HMC2_BLAGE	blattella g
11	980	39.0	461	1	HMCS_ARATH	P54873 arabidopsis
12	950.5	37.9	491	1	HMCS_YEAST	P54839 saccharomyc
13	900	35.8	447	1	HMCS_SCHPO	P54874 schizosacch
14	705	28.1	462	1	HMCS_CAEEL	P54871 caenorhabdi
15	249	9.9	163	1	HMCS_DICDI	P54872 dictyosteli
16	152	6.1	350	1	YD79_METKA	Q8tvl0 methanopyru
17	147	5.9	345	1	PKSG_BACSU	P40830 bacillus su
18	145.5	5.8	346	1	Y792_METTH	O26883 methanobact
19	144	5.7	350	1	Y677_PYRHO	Q51798 pyrococcus
20	138.5	5.5	350	1	Y972_PYRFU	pyrococcus
21	135	5.4	349	1	Y871_ARCFU	archaeoglob
22	132.5	5.3	343	1	YO15_ARCFU	O30256 archaeoglob
23	130	5.2	350	1	Y4EI_METAC	Q8zv4 methanosarc
24	127	5.1	349	1	YL85_PYRAE	pyrobaculum
25	126.5	5.0	350	1	Y4EL_METAC	Q8tiv0 methanosarc
26	123.5	4.9	716	1	YD69_PYRAB	Q9uy8 pyrococcus
27	123	4.9	351	1	Y132_THEVO	P2657 thermocoanaer
28	122	4.9	345	1	YF46_METJA	Q97cg9 thermoplasm
29	122	4.9	351	1	YE55_THEAC	Q58941 methanococc
30	121	4.8	580	1	NO56_MOUSE	Q9h187 thermoplasm
31	118.5	4.7	596	1	NO56_HUMAN	Q96621 mus musculu
32	116	4.6	348	1	YD49_SULTO	O00567 homo sapien
33	107.5	4.3	776	1	TOP1_RICPR	Q971k8 sulfolobus
34	106.5	4.2	1517	1	GLTB_ECOLI	P09831 escherichia
35	106	4.2	367	1	YQL2_CAEEL	Q09291 caenorhabdi
36	105	4.2	412	1	AATC_MOUSE	P05201 mus musculu
37	105	4.2	3358	1	PGCV_MOUSE	Q62059 mus musculu
38	103	4.1	489	1	YSV5_CAEEL	Q10011 caenorhabdi
39	103	4.1	1197	1	DPOM_PODAN	Q01529 podospora a
40	102	4.1	360	1	HIS8_LACLA	Q02135 lactococcus
41	102	4.1	1163	1	CQA_BACTF	Q9x597 bacillus th
42	101	4.0	365	1	YI73_AERPE	Q9ya80 aeropyrum p
43	100.5	4.0	776	1	TOP1_RICCN	Q92ih1 rickettsia
44	100	4.0	1136	1	C4BA_BACTI	P05519 bacillus th
45	99.5	4.0	494	1	ILVC_VIBPA	Q87cn4 vibrio para

ALIGNMENTS

```
RESULT 1
HMCS_HUMAN
ID   HMCS_HUMAN   STANDARD;   PRT;   520 AA.
AC   Q01581;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA
DE   synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HMGCS1 OR HMGCS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal adrenal gland;
RX   MEDLINE=94304197; PubMed=7913309;
RA   Rokosz L.L., Boulton D.A., Butkiewicz E.A., Sanyal G., Cueto M.A.,
RA   Lachance P.A., Hermes J.D.;
RT   "Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase:
RT   expression, purification, and characterization of recombinant
RT   wild-type and Cys129 mutant enzymes.";
RL   Arch. Biochem. Biophys. 312:1-13(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=93041939; PubMed=1358203;
RA   Russ A.P., Ruzicka V., Maerz W., Appelhans H., Gross W.;
RT   "Amplification and direct sequencing of a cDNA encoding human
RT   cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase.";
RL   Biochim. Biophys. Acta 1132:329-331(1992).
RN   [3]
RP   SEQUENCE FROM N.A., CHARACTERIZATION, AND MUTAGENESIS.
RC   TISSUE=Lung;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
```

```
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC       SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; X66435; CAA47061.1; -.
DR   EMBL; L25798; AAA62411.1; -.
DR   EMBL; BC000297; AAH00297.1; -.
DR   PIR; S27197; S27197.
DR   PIR; S45497; S45497.
DR   Genew; HGNC:5007; HMGCS1.
DR   MIM; 142940; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005625; C:soluble fraction; TAS.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; TAS.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt_1; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Transferase; Cholesterol biosynthesis; Multigene family.
FT   ACT_SITE    129    129
FT   MUTAGEN     129    129       C->A,S: LOSS OF ACTIVITY.
FT   CONFLICT    248    248       G->A (IN REF. 1).
FT   CONFLICT    251    251       K->N (IN REF. 1).
FT   CONFLICT    299    299       E->K (IN REF. 1).
FT   CONFLICT    364    364       Q->H (IN REF. 1).
FT   CONFLICT    420    420       P->Q (IN REF. 1).
FT   CONFLICT    519    520       EH->VW (IN REF. 1).
SQ   SEQUENCE    520 AA;  57293 MW;  C669212BF86CFF9B CRC64;
```

```
Query Match          98.8%;   Score 2480;   DB 1;   Length 520;
Best Local Similarity 91.9%;   Pred. No. 2.3e-173;
Matches 478;   Conservative 0;   Mismatches 0;   Indels 42;   Gaps 1;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT    60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI   120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-----------------------------   149
         |||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYAYGNARPTGGVGAVALL   180

Qy  150  ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI   198
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI   240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA   258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA   300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY   318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY   360

Qy  319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDDATPGSALDKITASLCDLKSRLDSRTGVAP   378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDDATPGSALDKITASLCDLKSRLDSRTGVAP   420
```

```
RESULT 2

ID   HMCS_RAT       STANDARD;   PRT;   520 AA.
AC   P17425;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA
DE   synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HMGCS1 OR HMGCS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90301491; PubMed=1972979;
RA   Ayte J., Gil-Gomez G., Hegardt F.G.;
RT   "Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-
RT   hydroxy-3-methylglutaryl coenzyme A synthase.";
RL   Nucleic Acids Res. 18:3642-3642(1990).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC       SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; X52625; CAA36852.1; -.
DR   PIR; S12736; S12736.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt_1; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Transferase; Cholesterol biosynthesis; Multigene family.
FT   ACT_SITE    129    129       POTENTIAL.
SQ   SEQUENCE    520 AA;  57433 MW;  CB2I3A27B0C177CB CRC64;
```

```
Query Match          94.1%;   Score 2362;   DB 1;   Length 520;
Best Local Similarity 87.5%;   Pred. No. 9.2e-165;
Matches 455;   Conservative 11;   Mismatches 12;   Indels 42;   Gaps 1;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT    60
         ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT    60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI   120
         |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCITVVQKLMERNSLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI   120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-----------------------------   149
         |||||||||||||||||||||||||||||
```

```
Qy  379  DVPAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG   438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  DVPAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG   480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH   478
         ||||||||||||||||||||||||||||||||||||||||
Db  481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH   520
```

Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYASGNARPTGGVGAVALL 180
Qy 150 ------------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPVIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYRKKI 240
Qy 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
Qy 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLIVSNQNGNMYTSSVYGSLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASAELFNQKTKASLLIVSNQNGNMYTSSVYGSLASVLAQY 360
Qy 319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP 420
Qy 379 DVFAENMKLREDTHHLVNYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHSLDEG 480
Qy 421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHSLDEG 480
Qy 439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db 439 DREDINSLCITVVQNLMERNSLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120
Db   61 DREDINSLCMTVVQNIMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Qy 481 VGLVHSNTATEHIPSPAKKVPRLPATSGEPESAVISNGEH 520

RESULT 3

ID   HMCS_CRIGR   STANDARD;   PRT;   520 AA.
AC   P13704;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA
DE   synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HMGCS1 OR HMGCS.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=86140166; PubMed=2869035;
RA   Gil G., Goldstein J.L., Slaughter C.A., Brown M.S.;
RT   "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the
RT   hamster. I. Isolation and sequencing of a full-length cDNA.";
RL   J. Biol. Chem. 261:3710-3716(1986).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC   SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   -!- ---
DR   EMBL; L00334; AAA37076.1; JOINED.
DR   EMBL; L00326; AAA37076.1; JOINED.
DR   EMBL; L00327; AAA37076.1; JOINED.
DR   EMBL; L00328; AAA37076.1; JOINED.
DR   EMBL; L00329; AAA37076.1; JOINED.
DR   EMBL; L00330; AAA37076.1; JOINED.
DR   EMBL; L00331; AAA37076.1; JOINED.
DR   EMBL; L00332; AAA37076.1; JOINED.
DR   EMBL; L00333; AAA37076.1; JOINED.
DR   PIR; A25332; A25332.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt_1; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
DR   Transferase; Cholesterol biosynthesis; Multigene family.
KW   Transferase; Cholesterol biosynthesis; Multigene family.
FT   ACT_SITE   129   129   POTENTIAL.
SQ   SEQUENCE   520 AA; 57318 MW; 2D4CAA87DEE5D6BB CRC64;

Query Match   93.7%; Score 2352; DB 1; Length 520;
Best Local Similarity   87.5%; Pred. No. 5e-164;
Matches 455; Conservative   9; Mismatches 14; Indels 42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT 60
Qy   61 DREDINSLCMTVVQNIMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db   61 DREDINSLCITVVQNLMERNSLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120
Qy 121 EGIDTTNACYGGTAAVFNAVNWIESSSWD----------------------------- 149
Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYATGNARPTGGVGAVALL 180
Qy 150 ------------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYRKKI 240
Qy 199 HAQWQKEGNDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Db 241 RAQWQKEGNDNDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300
Qy 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLIVSNQNGNMYTSSVYGSLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFNQKTKASLLIVSNQNGNMYTSSVYGSLASVLAQY 360
Qy 319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP 420
Qy 379 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHNLGDG 438
Db 421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHNLGDG 480
Qy 439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db 481 VGLVHSNTATEHIPSPAKKVPRLPATSGEPESAVISNGEH 520

RESULT 4

ID   HMCS_CHICK   STANDARD;   PRT;   522 AA.
AC   P23228;
DT   01-NOV-1991 (Rel. 20, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 2.3.3.10) (HMG-CoA
DE   synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HMGCS1 OR HMGCS.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Neoaves; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91112772; PubMed=1980405;
RA   Kattar-Cooley P.A., Wang H.-H.L., Mende-Mueller L.M., Miziorko H.M.;

```
RT   "Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes
RT   encode the cholesterogenic and ketogenic isozymes.";
RL   Arch. Biochem. Biophys. 283:523-529(1990).
RN   [2]
RP   CHARACTERIZATION, AND MUTAGENESIS.
RX   MEDLINE=95210320; PubMed=7696316;
RA   Misra I., Charlier H.A. Jr., Miziorko H.M.;
RT   "Avian cytosolic 3-hydroxy-3-methylglutaryl-CoA synthase: evaluation
RT   of the role of cysteines in reaction chemistry.";
RL   Biochim. Biophys. Acta 1247:253-259(1995).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC       SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M60657; AAA62737.1; -.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synt.
DR   Pfam; PF01154; HMG_CoA_synt_1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Transferase; Cholesterol biosynthesis; Multigene family.
FT   ACT_SITE    129    129       POTENTIAL.
SQ   SEQUENCE   522 AA;  57559 MW;  BFF7947C3E963C4C CRC64;

Query Match             83.5%;  Score 2097;  DB 1;  Length 522;
Best Local Similarity   76.6%;
Matches  400;  Conservative   36;  Mismatches   42;  Indels   44;  Gaps    3;

Qy          1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
              |||||| || |||||||||:||||||||||||||:||||||||||||||||||| ||||:
Db          1 MPGSLPVNTESCWPKDVGILVALEIYFPSQYVDQTELEKYDGVDAGKYTIGLGQSKMGFCS 60

Qy         61 DREDINSLCMTVVQNLMERNNISYDCIGRLEVGTEFIIDKSKSVKTNIMQLFEESGNTDI 120
              ||||||||| |||| |||||::|||||||||||||.|||||||||| ::||||||||||:
Db         61 DREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTETIIDKSKSVKTVLMQLFEESGNTDV 120

Qy        121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------ 149
              |||||||||||||| |||:|||||||||||
Db        121 EGIDTTNACYGGTAALFNAINWIESSSWDGRYALVVAGDIAVYATGNARPTGGAGAVAML 180

Qy        150 ------------GLRGTHMQHAYDFYKPDMLSEYPIYIDGKLSIQCYLSALDRCYSVYCKKI 198
                          ||||||||||||||||||:|||||:::|||||||||||||||||||| :||
Db        181 VGSNAPLIFERGLRGTHMQHAYDFYKPDMVSEYPVVDGKLSIQCYLSALDRCYSVYRNKI 240

Qy        199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRD-KNSIYSGLE 257
              ||||||||.|: |||||||||||||||||||||:.||:|:||||.||| :.  | |||||
Db        241 HAQWQKEGTDRGFTLNDFGFMIFHSPYCKLVQKSVARLLINDFLSDQNAETANGVFSGLE 300

Qy        258 AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQ 317
              || ||||||||||||||||||||| |||.||||||||||||||||||.||||.||:.|||
Db        301 AFRDVKLEDTYFDRDVEKAFMKASAELFNQKTKASLLVSNQNGNMYTPSVYGCLASILAQ 360

Qy        318 YSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSSALDKITASLCDLKSRLDSRTGVA 377
              ||| :||| ||  ||||||:||||||:.|||||||||||||||||||| ||:|||||  :|
Db        361 YSPEHLAGQRISEFSYGSGFAATLYSIRVTQDATPGSSALDKITASLSDLKARLDSRKCIA 420

Qy        378 PDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTNDDTLDE 437
              |||||||||:|::|||| ||||| |::.||||||||:||||||||||||||: .| ||:
Db        421 PDVFAENMKIRQETHHLANYIPQCSVEDLFEGTWYLLVRVDEKHRRTYARRPVMGDGPLEA 480

Qy        438 GVGLVHSNIATEHIPSPAKKVPRLPATA-AEPEAAVISNGEH 478
              || :||  |.|||||||||||||:|| : :.   .|:||| |
Db        481 GVEVVHPGIVHEHIPSPAKKVPRIPAITESEGVTVAISNGVH 522
```

```
Db          481 GVEVVHPGIVHEHIPSPAKKVPRIPAITESEGVTVAISNGVH 522

RESULT 5
ID   HMCM_HUMAN    STANDARD;    PRT;    508 AA.
AC   P54868;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE   (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE   A synthase).
GN   HMGCS2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95200282; PubMed=7893153;
RA   Mascaro C., Buesa C., Ortiz J.A., Haro D., Hegardt F.G.;
RT   "Molecular cloning and tissue expression of human mitochondrial 3-
RT   hydroxy-3-methylglutaryl-CoA synthase.";
RL   Arch. Biochem. Biophys. 317:385-390(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97449286; PubMed=9305755;
RA   Boukaftane Y., Mitchell G.A.;
RT   "Cloning and characterization of the human mitochondrial 3-hydroxy-3-
RT   methylglutaryl CoA synthase gene.";
RL   Gene 195:121-126(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gay L.J., Hulyk S.W.,
RA   Richards D.R., Worley K.C., Hale S., Garcia A.M., Gay L.J., Gibbs R.A.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 36-508 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=95154824; PubMed=7851882;
RA   Boukaftane Y., Duncan A., Wang S., Labuda D., Robert M.-F.,
RA   Sarrazin J., Schappert K., Mitchell G.A.;
RT   "Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic
RT   cloning, chromosome mapping to 1p12-p13, and possible role in
RT   vertebrate evolution.";
RL   Genomics 23:552-559(1994).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
```

CC   BODY BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- TISSUE SPECIFICITY: Liver and kidney.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   ---
CC   ---
DR   EMBL; X83618; CAA58593.1; -.
DR   EMBL; U81859; AAB72035.1; -.
DR   EMBL; U81851; AAB72036.1; -.
DR   EMBL; U81852; AAB72036.1; JOINED.
DR   EMBL; U81853; AAB72036.1; JOINED.
DR   EMBL; U81854; AAB72036.1; JOINED.
DR   EMBL; U81855; AAB72036.1; JOINED.
DR   EMBL; U81856; AAB72036.1; JOINED.
DR   EMBL; U81857; AAB72036.1; JOINED.
DR   EMBL; U81858; AAB72036.1; JOINED.
DR   EMBL; BC044217; AAH44217.1; -.
DR   EMBL; U12788; AAA92673.1; -.
DR   EMBL; U12789; AAA92674.1; -.
DR   PIR; S71623; S71623.
DR   Genew; HGNC:5008; HMGCS2.
MIM; 600234; -.
GO; GO:0005739; C:mitochondrion; TAS.
InterPro; IPR000590; HMG_CoA synt_AS.
InterPro; IPR008260; HMG_CoA_synth.
Pfam; PF01154; HMG_CoA synt; 1.
PROSITE; PS01226; HMG_COA SYNTHASE; 1.
Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
Multigene family.
TRANSIT   1   37   MITOCHONDRION (PROBABLE).
CHAIN   38   508   HYDROXYMETHYLGLUTARYL-COA SYNTHASE.
ACT_SITE 166   166   POTENTIAL.
SEQUENCE 508 AA; 56635 MW; BD362D631F7C3C80 CRC64;

Query Match 60.3%; Score 1513; DB 1; Length 508;
Best Local Similarity 60.4%; Pred. No. 7.5e-103;
Matches 282; Conservative 69; Mismatches 74; Indels 42; Gaps 1;

```
Qy    4 SLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCIDRE 63
          :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 AVPLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKYNNVEAGKYTVGLGQTRMGFCSVQE 100

Qy   64 DINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGI 123
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 DINSLCLTVVQRLMERIQLPWDSVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGI 160

Qy  124 DTTNACYGGTAAVFNAVNWIESSSWD---------------------- 149
          ||||||||||||||||||||||||
Db  161 DTTNACYGGTASLFNAANMESSSWDGRYAMVCGDIAVPSGNARPFGGAGAVAMLIGP 220

Qy  150 ------------GLRGTHMQHAYDFYKPNLASEYPIVDGKLSIQCYLSALDRCYSVYCKKHAQ 201
                      |||  |||||
Db  221 KAPLALERGLRGTHMENVYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYTSYRKKIQNQ 280

Qy  202 WQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEARGD 261
          :: :|::| |||  || | ||||||||||||||||||||||:||||||||||
Db  281 WKQAGSDRPFTLDDLQYMIFHTPFCKMVQKSLARLMFNDFLSASSDTQTSLYKGLEAPGG 340

Qy  262 VKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLIVSNQNGNMYTSSVVGSLASVLAQYSPQ 321
          :||||||||||| :||||| :||||||||||||||||||:||||||:|||||||||
Db  341 LKLEDTYTNKDLDKALLKASQMFDKKXTKASLYLSTHNGNMYTSSLYGCLASLLSHHSAQ 400

Qy  322 QLAGKRIGVFSYGSGLAAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVF 381
          :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  401 ELAGSRIGAFSYGSGLAASFFSFRVSQDAAPGSPLDKLVSSTSDLPKRLASRKCVSPEEF 460

Qy  382 AENNKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRP 428
          |||:|| : | | :| |||||||||| ||||||| |||||:||| ||||
Db  461 TEIMNQREQFYHKVNFSPPGDTNSLFPGTWYLERVDEQHRRKYARRP 507
```

RESULT 6

ID   HMCM_RAT   STANDARD;   PRT;   508 AA.
AC   P22791;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE   (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE   A synthase).
GN   HMGCS2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=90251660; PubMed=1971108;
RA   Ayte J., Gil-Gomez G., Haro D., Marrero P.F., Hegardt F.G.;
RT   "Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA
RT   synthases are encoded by two different genes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:3874-3878(1990).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC   BODY BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- TISSUE SPECIFICITY: Liver and kidney.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   ---
CC   ---
DR   EMBL; M33648; AAA41336.1; -.
DR   PIR; A35865; A35865.
InterPro; IPR000590; HMG_CoA synt_AS.
InterPro; IPR008260; HMG_CoA_synth.
Pfam; PF01154; HMG_CoA synt; 1.
PROSITE; PS01226; HMG_COA SYNTHASE; 1.
Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
Multigene family.
TRANSIT   1   37   MITOCHONDRION (PROBABLE).
CHAIN   38   508   HYDROXYMETHYLGLUTARYL-COA SYNTHASE.
ACT_SITE 166   166   POTENTIAL.
SEQUENCE 508 AA; 56912 MW; EC37693A5541D47E CRC64;

Query Match 59.5%; Score 1494; DB 1; Length 508;
Best Local Similarity 59.1%; Pred. No. 1.8e-101;
Matches 275; Conservative 76; Mismatches 72; Indels 42; Gaps 1;

```
Qy    6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCIDREDI 65
          ||| || ||||||||||||||||||||||||||||:||||||||||||:||||| |:|
Db   43 PLAKTDTWPKDVGILALEVYFPAQVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI 102

Qy   66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
          |||| |||| |||| :| ||| |||||||||||||||||:|||:||:||:||||||||||
Db  103 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 162

Qy  126 TNACYGGTAAVFNAVNWIESSSWD---------------------- 149
          ||||||||||||||||||||||||  ||
```

Db 163 TNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNRPFTGGAGAVAMLIGPKA 222
Qy 150 ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ 203
Db 223 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKIQNQWK 282
Qy 204 KEGNDKDFTLNDFGFMIFHSPYCKLVQKSIARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
Db 283 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSIARLMFNDFLSSSSDKQNNLYKGLEAFKGLK 342
Qy 264 LEDTYFDRDVEKAFMKKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQOL 323
Db 343 LEETYTNKDVDKALLKASLDMFNKKTKASLYLSTNNGNMYTSSLYGCIASLSHHSAQEL 402
Qy 324 AGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAE 383
Db 403 AGSRIGAFSYGSGLAASFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462
Qy 384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLIVRVDEKHRRTYARP 428
Db 463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRKYARRP 507

RESULT 7
HMCM_MOUSE
ID   HMCM_MOUSE   STANDARD;   PRT;   480 AA.
AC   P54869; Q64740;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE   (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE   A synthase) (Fragment).
GN   HMGCS2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHS; TISSUE=Liver;
RX   MEDLINE=95154824; PubMed=7851882;
RA   Boukaftane Y., Duncan A., Wang S., Labuda D., Robert M.-F.,
RA   Sarrazin J., Schappert K., Mitchell G.A.;
RT   "Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic
RT   cloning, chromosome mapping to 1p12-p13, and possible role in
RT   vertebrate evolution.";
RL   Genomics 23:552-559(1994).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC   BODY BIOSYNTHESIS
CC   -!- SUBCELLULAR LOCATION: Mitochondrial.
CC   -!- TISSUE SPECIFICITY: Liver and kidney.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.

DR   EMBL; U12790; AAA92675.1; -.
DR   EMBL; U12791; AAA92676.1; -.
DR   SWISS-2DPAGE; P54869; MOUSE.
DR   MGD; MGI:101939; Hmgcs2.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.

DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
KW   Multigene family.
FT   NON_TER   1   1
FT   TRANSIT   <1   9   MITOCHONDRION (PROBABLE).
FT   CHAIN   10   480   HYDROXYMETHYLGLUTARYL-COA SYNTHASE.
FT   ACT_SITE   138   138   POTENTIAL.
SQ   SEQUENCE   480 AA; 53786 MW; 7AA667A92862214D CRC64;

Query Match   59.5%; Score 1493; DB 1; Length 480;
Best Local Similarity   59.4%;
Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps   1;

Qy   6 PLNAEACWPKDVGIVALEIYPPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI 65
Db   15 PLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI 74
Qy   66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNIMQLFEESGNTDIEGIDT 125
Db   75 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 134
Qy 126 TNACYGGTAAVFNAVANWIESSSWD----------------------------------- 149
Db 135 TNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNARPTGGAGAVAMLIGPKA 194
Qy 150 ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ 203
Db 195 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKIQNQWK 254
Qy 204 KEGNDKDFTLNDFGFMIFHSPYCKLVQKSIARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
Db 255 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSIARLMFNDFLSSSSDKQNNLYKGLEAFRGLK 314
Qy 264 LEDTYFDRDVEKAFMKKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQOL 323
Db 315 LEETYTNKDVDKALLKASLDMFNQKTKASLYLSTNNGNMYTSSLYGCLASLLSHHSAQEL 374
Qy 324 AGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAE 383
Db 375 AGSRIGAFSYGSGLAASFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 434
Qy 384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLIVRVDEKHRRTYARP 428
Db 435 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRKYARCP 479

RESULT 8
HMCM_PIG
ID   HMCM_PIG   STANDARD;   PRT;   508 AA.
AC   O02734;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase, mitochondrial precursor
DE   (EC 2.3.3.10) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme
DE   A synthase).
GN   HMGCS2.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=97293201; PubMed=9164842;
RA   Adams S.H., Alho C.S., Asins G., Hegardt F.G., Marrero P.F.;
RT   "Gene expression of mitochondrial 3-hydroxy-3-methylglutaryl-CoA
RT   synthase in a poorly ketogenic mammal: effect of starvation during
RT   the neonatal period of the piglet.";
RL   Biochem. J. 324:65-73(1997).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase.
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-

Left column

```
CC   -!-  hydroxy-3-methylglutaryl-CoA + CoA.
CC        PATHWAY: RESPONSIBLE, TOGETHER WITH HMG-COA LYASE FOR KETONE
CC        BODY BIOSYNTHESIS.
CC   -!-  SUBCELLULAR LOCATION: Mitochondrial.
CC   -!-  SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; U90884; AAC48727.1; -.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Cholesterol biosynthesis; Mitochondrion; Transit peptide;
KW   Multigene family.
FT   TRANSIT       1     37       MITOCHONDRION (PROBABLE).
FT   CHAIN        38    508       HYDROXYMETHYLGLUTARYL-COA SYNTHASE.
FT   ACT_SITE    166    166       POTENTIAL.
SQ   SEQUENCE   508 AA;  56933 MW;  5479DE6F70B3C0F6 CRC64;

Query Match     Best Local Similarity   59.8%;
                Score 1466;  DB 1;  Length 508;  Pred. No. 2e-99;
Matches  278;  Conservative  68;  Mismatches  77;  Indels  42;  Gaps  1;

Qy    4 SLPLNAEACWPKDVGIVALEIYFPSQYDVQDQAELEKYDGVDAGKYTIGLGQAKMGFCTDRE 63
          ::|: |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 AVPVAKADTWPKDVGILALEVYFPAQYVDQTDLEKFDNVEAGRYTVGLGQTHMGFCSVQE 100

Qy   64 DINSLCMTVVQNLMERNNLSYDCIGRLEVGTEVGTEIIIDKSKSVKTNIMQLFEESGNTDIEGI 123
         |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 DINSLCITVVQRLMERTQLPWDSVGWLEVGTETIIDKSKSVKVTVLMELFQDSGNTDIEGI 160

Qy  124 DTTNACYGGTAAVFNAVNWIESSSWD------ 149
         |||||||||||   ||||:|:
Db  161 DTTNACYGGTASLFNAANWVESSAWDGRYAVVVCGDIAVYPRGNSRPTGGAGAVAMLVGP 220

Qy  150 -------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVVCKKHAQ 201
         ::  |||||||||||||||||||:|:||:|||||||||||||:|||
Db  221 EAPLALERGLRGTHMENAYDFYKPNATSEYPLVDGKLSIQCYLRALDRCYTLYRQKIEKQ 280

Qy  202 WQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEARGD 261
         |||||| ||||||||||||||||||:|::|||||||||||||:|||:|||||||
Db  281 WKQAGIERHFTLDDLQFMIFHTPFCKLVQKSLARLMFSDFLLADSDTQSSLYKGLEAFRG 340

Qy  262 VKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGGSLASVLAQYSPQ 321
         |||||| ||||  :|||||||||||||||||:||:||||||||||||||||||||||
Db  341 QKLEDTYANKDIEKAFQKASLDLFNKKTKPSLIYLSLHNGNMYTSSLYGCLASLLSQCSAQ 400

Qy  322 QLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVF 381
         |||||:|||||||||||| :::||||||||||||||||||:||:|||||||||::||
Db  401 DLAGSRIGAFSYGSGLAASFYSLRVSQDASPGSPLEKLVSSVSDLPERLASRKRVSPEEF 460

Qy  382 AENMKLREDTHHLLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYAR 426
         ||| :|||||||||||||||:|||||| |||||:|||:| |||||
Db  461 TEIMNQREQYYHKVNFTPPGDPNSLFPGTWYLERVDELYRRKYAR 505


RESULT 9
HMC1_BLAGE
ID   HMC1_BLAGE    STANDARD;    PRT;   453 AA.
AC   P54961;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase 1 (EC 2.3.3.10) (HMG-CoA synthase
DE   1) (3-hydroxy-3-methylglutaryl-coenzyme A synthase 1).
GN   HMGCS-1.
```

Right column

```
OS   Blattella germanica (German cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;
OC   Blattellidae; Blattellinae; Blattella.
OX   NCBI_TaxID=6973;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94039108; PubMed=7901012;
RA   Martinez-Gonzalez J., Buesa C., Piulachs M.D., Belles X.,
RA   Hegardt F.G.;
RT   "3-hydroxy-3-methylglutaryl-coenzyme-A synthase from Blattella
RT   germanica. Cloning, expression, developmental pattern and tissue
RT   expression.";
RL   Eur. J. Biochem. 217:691-699(1993).
```

```
CC   -!-  FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC        form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC        similarity).
CC   -!-  CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC        hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!-  PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-CoA PRIOR TO THE
CC        SYNTHESIS OF STEROLS AND ISOPRENOIDS SUCH AS JUVENILE HORMONES.
CC   -!-  SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   --------------------------------------------------------------------------
```

```
DR   EMBL; X73679; CAA52032.1; -.
DR   PIR; S38986; S38986.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Sterol biosynthesis; Multigene family.
FT   ACT_SITE    116    116       POTENTIAL.
SQ   SEQUENCE   453 AA;  50332 MW;  C79EB276270F348 CRC64;

Query Match     Best Local Similarity   56.6%;
                Score 1352.5;  DB 1;  Length 453;  Pred. No. 3.3e-91;
Matches  259;  Conservative  66;  Mismatches  84;  Indels  49;  Gaps  4;

Qy   13 WPKDVGIVALEIYFPSQYDVQDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTV 72
         ||:|||||||| |:||||| ||| |||| | :||||||:|||||:||||||||| |||||||
Db    2 WPSDVGIVALELIFPSQYVDQVDLEVYDNVSAGKYTVGLGQARMGFCTDRDINSLCLTV 61

Qy   73 VQNLMERNNLSYDCIGRLEVGTEVGTEIIIDKSKSVKTNIMQLFEESGNTDIEGIDTTNACYGG 132
         |::||||| : : ||||||||||||: ||||||||||:|:|:|||| : ||||||:||||||||
Db   62 VSRLMERWSIPYSQIGRLEVGTETILDKSKSVKVTVLMQLFKD--NTDIEGVDTTNACYGG 119

Qy  133 TAAVFNAVNWIESSSWD------ 150
         |:::||| :|:||||||
Db  120 TSALFNAISWVESSSWDGRYALVVAGDIAVYAKGSARPTGGAGAVAMLVGANAPLVFDRG 179

Qy  151 LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVVCKKIHAQWQKEGNDKD 210
         :||:||||||||||||  | :|| ||||||||||||||:|| : | ::|
Db  180 VRSSHMQHAYDFYKPDLSSLYPTVDGKLSIQCYLSALDHCYQLYCSKI----QKOLGEK- 234

Qy  211 FTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEARGDVKLEDTYFD 270
         |::: : :::||:||||||||||||:|:|||: : : :  :: ||||:|||||||||||
Db  235 FDIERLDAVLFHAPYCKLVQKSLARLVLNDFVRASEEERTTKYSSLEALRGVKLEDTYFD 294

Qy  271 RDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGGSLASVLAQYSPQQLAGKRIGV 330
         |:||||.|  :.  ||:|||.||::.|||.||||.|:|||.|:.:. .::|:||||||:.:
Db  295 REVEKAVMTYSKNMFEEKTKPSLILLANQVGNMYTPSLYGGLSVLIVSKSAQELAGKRVAL 354

Qy  331 FSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRED 390
         |||| ||| : :||:::.|:: |:|:::::.|| .:.::|:. :::.||::
Db  355 FSYGGGLASSMFSLRISSDSAKSSLQRLVSNLSHIKPQLDLRHKVSPEEPAQTMETREH 414
```

Qy 391 THHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRP 428
　　　　　 |||| ||| ||||| ||||||| ||||| :| || |
Db 415 NHHKAPYTPEGSIDVLFPGTWYLESSVDSIYRSYKQVP 452

Db 182 GVRASHMKHAYDFYKPDMLSEYPVVDGKLSVQCYLSALDHCYPRFCSKTEKYLKRCGKEN 241
Qy 210 -DFILNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTY 268
Db 242 TKIDLDYFDAFVFHSPYCKLVQKSVARLVLNDFI--QYPEK---YQDLQQLRNLKFEDTY 296

RESULT 10
HMC2_BLAGE

ID   HMC2_BLAGE   STANDARD;   PRT;   455 AA.
AC   P54870;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase 2 (EC 2.3.3.10) (HMG-CoA synthase
DE   2) (3-hydroxy-3-methylglutaryl coenzyme A synthase 2).
GN   HMGCS-2.
OS   Blattella germanica (German cockroach).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Orthopteroidea; Dictyoptera; Blattaria; Blaberoidea;
OC   Blattellidae; Blattellinae; Blattella.
OX   NCBI_TaxID=6973;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94216267; PubMed=7909314;
RA   Buesa C., Martinez-Gonzalez J., Casals N., Haro D., Piulachs M.D.,
RA   Belles X., Hegardt F.G.;
RT   "Blattella germanica has two HMG-CoA synthase genes. Both are
RT   regulated in the ovary during the gonadotrophic cycle.";
RL   J. Biol. Chem. 269:11707-11713(1994).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC   similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-COA PRIOR TO THE
CC   SYNTHESIS OF STEROLS AND ISOPRENOIDS SUCH AS JUVENILE HORMONES.
CC   -!- SIMILARITY: Belongs to the HMG-COA synthase family.

DR   EMBL; X75516; CAA53652.1; -.
DR   PIR; A53565; A53565.
DR   InterPro; IPR000590; HMG_CoA_synt AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Transferase; Sterol biosynthesis; Multigene family.
FT   ACT_SITE   120   120   POTENTIAL.
SQ   SEQUENCE   455 AA;   51425 MW;   DD030356548A63B1 CRC64;

Query Match   50.7%;   Score 1272.5;   DB 1;   Length 455;
Best Local Similarity   53.7%;   Pred. No.2.2e-85;
Matches 246;   Conservative   64;   Mismatches 101;   Indels 47;   Gaps   4;

Qy   11 ACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCM   70
　　　　　 | ||:|||| :|::|||| |||||| ||: | |||:|||||:|||:|||||||||||
Db   2 AHWPEDVGIGIEMIFPSLVDQAELETYDEVSPGKXYTMGLGQDKMGVCTDREDINSLCL   61

Qy   71 TVVQNLMERRNLSYDCIGRLEVGTETIIDKSKSVKNLMQLFEESGNTDIEGIDTTNACY 130
　　　　　 |.|.:|||| |: |: || :||||||||::|||||:::|||||||||||:|||||:||||
Db   62 TAVDKLMERNNIDYNDIGWLEVGTETILDKVKSVKITVLMQLFEESGNTDVEGIDTINACY 121

Qy 131 GGTAAVFNAVNWIESSSWD------ 149
　　　　　 ||||| |||:.|||||||||
Db 122 RGTAALFNALIWIESSSWDGRYAIVVAADIAIYAKECSPTGGAGALLMLIGANAPIVIDR 181

Qy 150 GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSICYLSALDRCYSVVCKKIHAQWQKEGNDK 209
　　　　　 || |:|:|||||||||||||||| | |||:||||||| :|:||::| : :
Db 182 ...

Qy 269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSGLASVLAQYSPQQLAGKRI 328
Db 297 FDRIEKIFMDKSSKQLPEKKTKPSLMLANQVGNMYTTSLYGGLVSLLISEDIGELAGKCI 356

Qy 329 GVFSYGSGLAATLYSLKVTQDATPGSSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
Db 357 CMFSSYGSGFAASMFSLHISTDSSPGSTLSRLVTNLTHIKPQVQORVKLLSPGEFFNIMEIR 416

Qy 389 EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYAR 426
Db 417 EQNHKAPYTPVASPNTLFPGTWYLESIDSMHRRKYKR 454

RESULT 11
HMCS_ARATH

ID   HMCS_ARATH   STANDARD;   PRT;   461 AA.
AC   P54873; Q8L721; Q9S707;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE   (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HMGS OR MVA1 OR At4g11820 OR T26M18.30.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=cv. Landsberg erecta;
RX   MEDLINE=96144274; PubMed=8566777;
RA   Montamat F., Guilloton M., Karst F., Delrot S.;
RT   "Isolation and characterization of a cDNA encoding Arabidopsis
RT   thaliana 3-hydroxy-3-methylglutaryl-coenzyme A synthase.";
RL   Gene 167:197-201(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Connolly E.L., Learned R.M.;
RT   "Post-transcriptional regulation of HMG-CoA synthase expression in
RT   Arabidopsis thaliana.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueten J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,

RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Cooke R., Berger C., Montfort A., Casacuberta E.,
RA   Chefdor F., Barber P., Berger C.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M.A., Martiensen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=cv. Columbia;
RA   Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT   "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT   SSP consortium (Salk/Stanford/PGEC).";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC   similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC   synthesis of sterols and isoprenoids.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=P54873-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=P54873-2; Sequence=VSP_008902;
CC   -!-
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
DR   EMBL; X83882; CAA58763.1; -.
DR   EMBL; U79160; AAD00297.1; -.
DR   EMBL; U79161; AAD00298.1; -.
DR   EMBL; AL078606; CAB44320.1; -.
DR   EMBL; AL161532; CAB78225.1; -.
DR   EMBL; AY140008; AAM98150.1; -.
DR   EMBL; BT008492; AAP37851.1; -.
DR   PIR; T09341; T09341.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Sterol biosynthesis; Transferase; Alternative splicing.
FT   ACT_SITE   117   117
FT   VARSPLIC   1   55   Missing (in isoform 2).
FT   /FTId=VSP_008902.
FT   POTENTIAL.
FT   CONFLICT   306   306   A -> S (IN REF. 1).
FT   CONFLICT   342   342   K -> N (IN REF. 1).

RESULT 12
HMCS_YEAST

ID   HMCS_YEAST   STANDARD;   PRT;   491 AA.
AC   P54839;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE   (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   ERG13 OR HMGS OR YML126C OR YM4987.09C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 28383 / FL100;
RA   Kribii R., Cordier H., Karst F.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC   form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC   similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-

SQ   SEQUENCE   461 AA; 51094 MW; F44908369AFBC5A8 CRC64;
     Query Match   39.0%; Score 980; DB 1; Length 461;
     Best Local Similarity   43.3%; Pred. No. 4.9e-64;
     Matches 198; Conservative 79; Mismatches 128; Indels 52; Gaps   6;

```
Qy         15 KDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
              :||||:||:::|||:|:::|:|  |||||| ||||||||||||||||:|:|||:|||:||
Db          3 KNVGILAMDIYFPPTCVQQEALEAHDGASKGKYTIGLGQDCLAFCTELEDVISMSPNAVT 62

Qy         75 NLMERNNLSVDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNGNTDIEGIDTTNACYGGTA 134
              :|::||||::|:|||||||:|||||||:|||||||||||||:|:|||:|||::|||||||
Db         63 SLFFKYKIDDNQIGRLEVGSETVIDKSKSIKTFLMQLFEKCGNTDVEGVDSTNACYGGTA 122

Qy        135 AVFNAVNWIESSSWDG------------------------------------------- 152
              :  |:|:||                                                LR
Db        123 ALINCVNWVESNSWDGRYGLVICTDSAVYAEGPARPTGGAAAIAMLIGPDAPIVFESKLR 182

Qy        153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCSVYCKKIHAQWMQKEGNDKDFT 212
              |:|||||||  |  ||||||||||:||:||||:|||||||:||||||||   :  |:||:
Db        183 ASHMAHVYDFYKPNLASEYPVVDGKLSQTCYLMALDSCYKHLCNKFE---KIEG--KBFS 237

Qy        213 LNDFGFMIFHSPYCKLVQKSIARMLINDFLNDQNRDKNSIYSGLEARFGDVKLEDTYFDRD 272
              |:|  |:|:|||||:|||:|||||:|||:|  |:| :  ||::|||||  | | | |:|
Db        238 INDADYIVFHSPYNKLVQKSFARLLYNDFLRNASSIDEAAKEKFTPYSSLTLDESYQSRD 297

Qy        273 VEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFS 332
              |:|||:|||:|||:||  |:|||:  ||  |||||||||||||||||||| |:||:||||
Db        298 LEKVSQQIAKPFYDAKVQPTTLIPKEVGNMYTASLYAAFASLIHK-KHNDLAGKRVVMFS 356

Qy        333 YGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTH 392
              |||||||:||:|||:|  |||:||||||:||||||||| |:||||||:||:||||| |
Db        357 YGSGSTAMFSLRLNDNKPPFSISN--IASVMDVGGKLKARHEYAPEKFVETMKLMEHRY 414

Qy        393 HLVNYI--PQGSIDSLFEGTWYLVRVDEKHRRTYARR 427
              ||  |:   ||| || |:  ||||||||:|||| | |
Db        415 GAKDFVTTKEGIIDLLAPGIYYLKEVDSLYRRFYGKK 451
```

```
CC   hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC       synthesis of sterols and isoprenoids.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X96617; CAA65437.1; -.
DR   EMBL; Z50178; CAA90557.1; -.
DR   PIR; S58202; S58202.
DR   GermOnline; 142661; -.
SGD; S0004595; ERG13.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Sterol biosynthesis.
FT   ACT_SITE    159    159
SQ   SEQUENCE   491 AA;  55013 MW;  44DFF3COB035609E CRC64;

Query Match            37.9%;  Score 950.5;  DB 1;  Length 491;
Best Local Similarity  43.4%;  Pred. No. 7;7e-62;
Matches  198;  Conservative  72;  Mismatches  127;  Indels  59;  Gaps  7;

Qy     14  PKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV  73
               ::|||..:|||:.||.|...||.:.||||.::|||||..::|..||.:.:||:.|:||.
Db     14  PKDVGILALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV  73

Qy     46  PQNVGIKGIQIYIPTQCVNQSELEKFDGVSQGKYTIGLGQTNMSFVNDREDIYSMSLTVL  105
Db     46  ...

Qy     74  QNLMERNNLSYDCIGRLEVGTETIIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGT  133
               :||:|||:.:||||.|.|||.|||||||||.||.:|||.:||....:.:|||:||||||||
Db    106  SKLIKSYNIDTNKIGRLEVGTETLIDKSKSVKSVLMQLFGE--NTDVEGIDTLNACYGGT  163

Qy    134  AAVFNAVAVWIESSSWDG-                                           LR  152
               :||||||||||.||.:||.                                           :|
Db    164  NALFNSLNWIESNAWDGRDAIVVCGDIAIYDKGAARPTGGAGTVAWWIGPDAPIVPDSVR  223

Qy    153  GTHMQHAYDPYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQW---QXBGNDK  209
Db    224  ASYMEHAYDPYKPDFTSEYPYVDGHFSLTCYVKALDQVVKSYSKKAISKGLVSDPAGSDA  283

Qy    210  DFTLNDFGFMIFHSPYCKLVQKSIARMLLNDFLNDQNRDKNSIYSGLRA-FGDVKLEDTY  268
Db    284  LNVLKYFDYNVFHVPTCKLVTKSYGRLLYNDF----RANPQLFPEVDAELATRDYDESL  338

Qy    269  FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNNMYTSSYYGSLASVLAQYSFPCQLAGKRI  328
Db    339  TDKNIEKTFVNVAKPFHKERVAQSLIVPTNTGNMYTASVYAAFASLLNYVGSDDLQGKRV  398

Qy    329  GVFSYGSGLAATLYSLKVTQDAIPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR  388
Db    399  GLFSYGSGLAASLYSCKIVGD------VQHIKEL-DITNKLAKRIETPDKDYEAAIELR  451

Qy    389  EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTY  424
Db    452  ENAHLKKNFKPQGSIEHLQSGVVYLTNIDDKFRRSY  487
```

```
OS   Schizosaccharomyces pombe (fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetidae; Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96535436; PubMed=8750242;
RA   Katayama S., Adachi N., Takao K., Nakagawa T., Matsuda H.,
RA   Kawamukai M.;
RT   "Molecular cloning and sequencing of the hcs gene, which encodes 3-
RT   hydroxy-3-methylglutaryl coenzyme A synthase of Schizosaccharomyces
RT   pombe.";
RL   Yeast 11:1533-1537(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC       synthesis of sterols and isoprenoids.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U32187; AAB17601.1; -.
DR   EMBL; Z98530; CAB11060.1; -.
DR   PIR; S61875; S61875.
DR   GeneDB_SPombe; SPAC4F8.14c; -.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Sterol biosynthesis.
FT   ACT_SITE    118    118
SQ   SEQUENCE   447 AA;  49239 MW;  919BDDBD9207B886 CRC64;
```

```
RESULT 13
ID   HMCS_SCHPO     STANDARD;     PRT;   447 AA.
AC   P54874;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE   (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   HCS OR SPAC4F8.14C.
```

RESULT 14

```
ID   HMCS_CAEEL      STANDARD;      PRT;   462 AA.
AC   P54871; Q22962;
DT   01-OCT-1996 (Rel. 34, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE   (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN   F25B4.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Gattung S.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 5-462 FROM N.A.
RA   Mitchell G.A.;
RL   Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC       synthesis of sterols and isoprenoids.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U64842; AAB37084.1; -.
DR   EMBL; U12787; AAA22672.1; -.
DR   PIR; T25726; T25726.
DR   WormPep; F25B4.6; CE09624.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Transferase; Sterol biosynthesis.
FT   ACT_SITE    124                   POTENTIAL.
SQ   SEQUENCE   462 AA;   51415 MW;  89BDE3825B88F6D9F CRC64;
```

Query Match   35.8%; Score 900; DB 1; Length 447;
Best Local Similarity   41.5%; Pred. No. 3.2e-58;
Matches 191; Conservative 72; Mismatches 133; Indels 64; Gaps 6;

```
Qy       15 KDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
            |||| ||:|||:|| |||.:..|| || : |.: :|| | :|||||| |: |.|:::|:||
Db        6 KDIGIGLVLYTPNQYVEEQAALEAHDGVSTGKYTIIGLGLTKMAFVDREDIYSFGLTALS 65

Qy       75 NLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGTA 134
            |||||||||||||||||| |||||||||||||||||||||| ||||||||||||| |:||
Db        66 QLIKRYQIDISKGIGRLEVGTETIIDKSKSVKSVLMQLFGD--NHNVEGIDCVNACYGGVN 123

Qy      135 AVFNAVNWIESSSWD----------------------------------------GLR 152
            :||.||| ::|:.||                                       |||
Db       124 ALFNIDWIESSAAWDGRDGIVVAGDIALYAKGNARPTGGAGCVALLVGPNAPIVFEPGLR 183

Qy      153 GTHMQHAYDFYKPEDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFT 212
            :||:||||||| | |||||||||||| |||||||||||||:||.||||||||||||||| |
Db       184 GTYMQHAYDFYKPDLTSEYPYVDGHFSLECYVKALDGYAYANNVRDVA---KNGKSQGLG 240

Qy      213 LNDFGFMIFHSPYCKLVQKSLARMLLNDFL----NDQNRDKNSIYSGLEAFGDVKLEDTY 268
            |||||||||||||||||||||||||||||.    |||||||||||||||||||:||||:|
Db       241 LDRFDYCIFHAPYCKQVQKAYARLLYTDSAAEPSNPELEGVRELLSTLDA-----KKSL 294

Qy      269 FDRDVEKAFMKASSELFSQKTKASLLIVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRI 328
            ||||||||||||||||||||||||||.||:||||||||||||||||||||||||||||||:
Db       295 TDKALEKGLMAITKERFNRVSPSVYAPTNCGNMYTASIFSCLTALLSRVPADELKGKRV 354

Qy      329 GVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||:|||||:
Db       355 GAVSYGSGLAASFFSFVVKGDVSE------IAKKTNLVNDLDNRHCLTPTQYEEAIELR 407

Qy      389 EDTHHLVNYIPQGSIDSLFEGTWYLLVRVDEKHRRTYARRP 428
            : :|||||||||||||||||||||||||||||||:|||:.
Db       408 HQAHLKKNFTPKGSIERLRSGTYVLTGIDDMFRRSYSVKP 447
```

RESULT 15

```
ID   HMCS_DICDI      STANDARD;      PRT;   163 AA.
AC   P54872;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase (EC 2.3.3.10) (HMG-CoA synthase)
DE   (3-hydroxy-3-methylglutaryl coenzyme A synthase) (Fragment).
GN   HGSA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Dhar M.S., Hauser L.J., Olins D.E., Olins A.L.;
```

Query Match   28.1%; Score 705; DB 1; Length 462;
Best Local Similarity   37.1%; Pred. No. 5.7e-44;
Matches 176; Conservative 77; Mismatches 151; Indels 70; Gaps 13;

```
Qy       16 DVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQN 75
            ||| |:|:||||||||||||| || || :.: |||.|||||| |: |.::|||||:||::|
Db        13 DVGIGAIELYFPQNFVDQNDLEKFNNVSSGKYTIGLGQQOMGFCSDNEDIVSISLRTVTRK 72

Qy       76 LMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGTAA 135
            ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||:|||
Db        73 LIETYKISTDSIGCLVVGTETMIDKSKSVKTALMDLF--PGNSDIEGVDIKNACFGGAQA 130

Qy      136 VFNAVNWIE--------------------------------------SSSWDGLRGT- 154
            |||||||:                                      |||||||||:
Db       131 LLHAIDWTVNHPLDKKNAIVVADIAIYEEGPARCTGGAGAIAFLICPDASIPIDRQFS 190

Qy      155 --HMQHAYDFYKP--DMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKD 210
              ||||||||||  |||||||||||||:|||||||||||||||||||||||||||||||
Db       191 ACHMQNTWDFFKPITPIPSSEYPVVDGSLSLSSSYLEAAVRMTYTYFISKVNR--HTTGID-- 246

Qy      211 FTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFD 270
            ||:|||| ||||||||:||||||||||||| ||||:||||||||||||||:||||||||
Db       247 -GLNSFDGVFLHSPTKMVQKGLAVM---NYTDSQLRHKQLNGNGV--DHKLDEN--D 296

Qy      271 RDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAG-KRIG 329
            ||:||||||||||||||||||:||..||:||||||:||||||||||||||||||| ||||
Db       297 RAGLAKMIELSAQVWKEKTDPYLVFNRRIGNMYTPSLFAQLLAYLA--ADDCVTGBKSIL 354

Qy      330 VFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRE 389
            ||||||||||:||||||||||||||||||||||| ||||||||:|||||||||||||||:
Db       355 FFAYGSGLASAIFGCRVRQT----SNLDKIRQVAIRAIRKLDDRIQFTPEEFTETLQKRE 410

Qy      390 DTHHLVNYIPQGSIDSLFEGTWYLLVRVDEKHRRTYARRPPNDDTLDEGVGLVH 443
            |:|||||||||||||||||||||||:|||||||:||:|::|||| |:||||:|
Db       411 VFLRSKEIPKSPSETSLFPNTYFLDNMDKLYRRSYTLHEEPNG--VQNGNGIHH 462
```

```
RT   "A putative partial cDNA clone of HMG-CoA synthase from Dictyostelium
RT   discoideum.";
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: This enzyme condenses acetyl-CoA with acetoacetyl-CoA to
CC       form HMG-CoA, which is the substrate for HMG-CoA reductase (By
CC       similarity).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + acetoacetyl-CoA = (S)-3-
CC       hydroxy-3-methylglutaryl-CoA + CoA.
CC   -!- PATHWAY: Production of mevalonate from HMG-CoA prior to the
CC       synthesis of sterols and isoprenoids.
CC   -!- SIMILARITY: Belongs to the HMG-CoA synthase family.
CC   -----------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; L24114; AAA91055.1; -.
DR   DictyBase; DDB0001904; hgsA.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; PARTIAL.
KW   Transferase; Sterol biosynthesis.
FT   NON_TER       1      1
SQ   SEQUENCE   163 AA;  18365 MW;  ED9BA773958A60C4 CRC64;
```

```
Query Match             9.9%;  Score 249;  DB 1;  Length 163;
Best Local Similarity   39.3%;  Pred. No. 2.4e-11;
Matches   57;  Conservative  26;  Mismatches  50;  Indels  12;  Gaps   4;

Qy   285  FSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSL  344
          ::  |::|| ||||: ||||:::|:|| :|||| :||: ||||||||||||:: :|
Db     6  YATKVAPTTLLAKQLGNTYCGSTYSGLLSLLIDEKS-NDLVGKRVLTFSYGSGLAASAFSF   64

Qy   345  KVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSID  404
          || :    |||:::|  |: |||        | ::||  ||||||||||:
Db    65  KVEKPI--NHIVEKV-----DLKNRLAKRVREPEIFTEKLSLRETRHNLKNYVPSDETT  117

Qy   405  SLFEGTWYLVRVDE----KHRRTYA  425
          ::|:|||||::||  ||: |||:
Db   118  NMFPGSFYLSSVDKAGIRKFPRTYS  142
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 05:08:06 ; Search time 762 Seconds
                   (without alignments)
                   11161.276 Million cell updates/sec

Title: US-10-622-516-1
Perfect score: 2002
Sequence: 1 cgcctcccagcgactctcgg.........aaaaaaaaaaaaaaaaaaa 2002
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 3373863 seqs, 212409041 residues
Total number of hits satisfying chosen parameters: 6747726
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
               1: geneseqn19108:*
               2: geneseqn1990s:*
               3: geneseqn2000s:*
               4: geneseqn2001as:*
               5: geneseqn2001bs:*
               6: geneseqn2002s:*
               7: geneseqn2003as:*
               8: geneseqn2003bs:*
               9: geneseqn2003cs:*
               10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	2002	6	ABS54409	Human hyd
2	2002	100.0	2002	7	ABX93299	cDNA enco
3	1693	84.6	3722	10	ADE76933	Human CDN
4	1605.2	80.2	2540	9	AAS84743	DNA encod
5	1370	68.4	1650	7	ACC62336	Human NOV
6	1370	68.4	1650	7	ACC62328	Human NOV
7	1370	68.4	1650	7	ACC62326	Human NOV
8	1370	68.4	1650	7	ACC62330	Human NOV
9	1370	68.4	1650	7	ACC62338	Human NOV
10	1370	68.4	1650	7	ACC62334	Human NOV
11	1370	68.4	1650	7	ACC62327	Human NOV
12	1370	68.4	1650	7	ACC62332	Human NOV
13	1370	68.4	3008	4	ADE38394	Human pro
14	1322.4	66.1	2008	4	AAH34834	Human col
15	1302	65.0	1564	7	ACC62337	Human NOV
16	1302	65.0	1564	7	ACC62339	Human NOV
17	1299.4	64.9	1601	7	ACC62331	Human NOV
18	1299	64.9	1593	7	ACC62329	Human NOV
19	1299	64.9	1612	7	ACC62340	Human NOV
20	1299	64.9	1612	7	ACC62335	Human NOV
21	1295.4	64.7	1608	7	ACC62341	Human NOV
22	1295.4	64.7	1608	7	ACC62333	Human NOV
23	1288	64.3	1685	9	ADD14748	Human src
24	1180.6	59.0	3275	8	ACF25333	Rat cytos
25	1180.6	59.0	3275	9	ADB58295	Toxicity-
26	1180.6	59.0	3275	9	ADB52844	Primary r
27	794.6	39.7	1824	3	AAT89089	Avian 3-h
28	476	23.8	28001	6	ABS54410	Human hyd
29	476	23.8	28001	7	ABX93300	Gene enco
30	460.4	23.0	565	5	AAS61469	Lung smal
31	448.2	22.4	505	7	ABZ84645	Toxicolog
32	436.4	21.8	615	3	AAC79203	Human lun
33	436.4	21.8	615	5	AAD23279	Human lun
34	436.4	21.8	615	9	ADD66553	Human lun
35	436.4	21.8	615	9	ADE87807	Human lun
36	436.4	21.2	2093	9	AAC98123	Human col
37	436.4	21.1	2088	4	AAK52002	Human pol
38	421.8	21.1	2377	9	ADE53675	Human pro
39	421.8	21.0	2058	6	ABL69551	Prostate
40	420.2	21.0	2058	6	ABN97315	Gene #381
41	420.2	21.0	2058	7	ACC50182	Breast ca
42	420.2	21.0	2132	9	ADB75349	Prostate
43	420.2	21.0	2132	9	ADE09760	Novel DNA
44	418.2	20.9	1845	8	AAD57512	Human enz
45	413.2	20.6	1994	4	AAH22426	Rat mitoc

ALIGNMENTS

RESULT 1
ABS54409

```
ID   ABS54409 standard; cDNA; 2002 BP.
XX
AC   ABS54409;
XX
DT   22-NOV-2002  (first entry)
XX
DE   Human hydroxymethylglutaryl-coenzyme A synthase protein cDNA.
XX
KW   Human; gene; ss; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;
KW   HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;
KW   antibody; synthase; carcinoma.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   5'UTR           1..144
FT                   /*tag= a
FT   CDS             145..1581
FT                   /*tag= b
FT                   /product= "Hydroxymethylglutaryl-coenzyme A synthase"
FT   variation       replace(717,T)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1582..2002
FT                   /*tag= d
XX
PN   US6436692-B1.
XX
PD   20-AUG-2002.
XX
PF   29-MAR-2001; 2001US-00819993.
XX
PR   29-MAR-2001; 2001US-00819993.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-689940/74.
DR   P-PSDB; ABG32726.
XX
PT   New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA
PT   synthase, useful as model for the development of human therapeutic
PT   targets and for identifying therapeutic proteins.
```

XX
PS   Claim 1; Fig 1; 62pp; English.
XX
CC   The invention discloses an isolated nucleic acid molecule encoding a
CC   human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC   important in cholesterologenesis. The polynucleotide and polypeptide are
CC   useful as models for the development of human therapeutic targets, to aid
CC   in the identification of therapeutic proteins and as targets for the
CC   development of human therapeutic agents that modulate the activity of the
CC   polypeptide in cells and tissues. The polynucleotide is useful for
CC   monitoring the effectiveness of modulating compounds on the expression or
CC   activity of the enzyme gene in clinical trials and in treatment, in
CC   diagnostic assays for qualitative changes in expression of enzyme nucleic
CC   acid, to detect mutations in enzyme genes and gene expression products,
CC   such a mRNA, for testing an individual for a genotype and as a diagnostic
CC   target that can be used to tailor treatment in an individual. The
CC   polypeptide is useful to raise antibodies, as a reagent in assays
CC   designed to quantitatively determine levels of the protein in biological
CC   fluids, as markers for tissues in which the corresponding protein is
CC   preferentially expressed (e.g. carcinomas), in biological assays related
CC   to members of the synthase subfamily, in drug screening assays, in
CC   competition binding assays, in cell-based or cell-free systems, in
CC   pharmacogenomic analysis and for treating a disorder characterised by an
CC   absence of, inappropriate, or unwanted expression of the polypeptide. The
CC   sequence presented is the human HMG-CoA synthase protein cDNA, the gene
CC   for which is located on chromosome 5
XX
SQ   Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;

Query Match   100.0%; Score 2002; DB 6; Length 2002;
Best Local Similarity   100.0%;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 CGCCTCCCAGCGACTCTCGGCAGTGCCGGAGTCGGGGTTGGGCGGCTATAAAGCTGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CGCCTCCCAGCGACTCTCGGCAGTGCCGGAGTCGGGGTTGGGCGGCTATAAAGCTGGT 60

Qy   61 AGCGAAGGGGAGGCGGCGCCCGACGTGTCCTTTCGTGGCTCACTCCCTTTCCCTGCGCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGCGAAGGGGAGGCGGCGCCCGACGTGTCCTTTCGTGGCTCACTCCCTTTCCCTGCGCCG 120

Qy   121 CTCGGTCACGCTTGCTCTTTCACCATGCCCTGGATCACTTCCTTTGAATGCAGAAGCTTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTCGGTCACGCTTGCTCTTTCACCATGCCCTGGATCACTTCCTTTGAATGCAGAAGCTTGC 180

Qy   181 TGGGCCAAAGATGTGGGAATTGTTGCCCTTGAGATCTTATTTTCCTTCTCAATATGTTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGGCCAAAGATGTGGGAATTGTTGCCCTTGAGATCTTATTTTCCTTCTCAATATGTTGAT 240

Qy   241 CAAGCAGAGCAGTTGGAAAAATATGATGGTGTAGACTGGGAAAAGTATACCAATTGGCTTGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAAGCAGAGCAGTTGGAAAAATATGATGGTGTAGACTGGGAAAAGTATACCAATTGGCTTGGGC 300

Qy   301 CAGGCCAAGATGGGCTTCGCACAGATAGAGAAGATATTAACTCTCCTTTGCATGACTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGCCAAGATGGGCTTCGCACAGATAGAGAAGATATTAACTCTCCTTTGCATGACTGTG 360

Qy   361 GTTCAGAAATCTTATGGAGAGAGAAATAACCTTTCCTATGATTGCCATTGGCAGCTGTTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTTCAGAAATCTTATGGAGAGAGAAATAACCTTTCCTATGATTGCCATTGGCAGCTGTTT 420

Qy   421 GGAACAGAGACAAATCATCGACAAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGGAGGC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGAACAGAGACAAATCATCGACAAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGGAGGC 480

Qy   481 GAAGAGTCTCGGGAATACAGATATAGAAGGAATCGACACAACTAATTGCATGGCTGTCGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAAGAGTCTCGGGAATACAGATATAGAAGGAATCGACACAACTAATTGCATGGCTGTCGGT 540

Qy   541 ACAGCTGCTGTCTTCAAATGCTGTTAACTGGATTGAGTGTCCAGCTCTTGGGATGGGCTTCGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACAGCTGCTGTCTTCAAATGCTGTTAACTGGATTGAGTGTCCAGCTCTTGGGATGGGCTTCGT 600

Qy   601 GGGACACATATGCCAACATGCCTATGAGTTGATTTTTTACAAGCCTGATATGCTTGATCTCT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGGACACATATGCCAACATGCCTATGAGTTGATTTTTTACAAGCCTGATATGCTTGATCTCT 660

Qy   661 ATAGTTAGATATGGAAAAACTTCCATACAGTCGTACCTCTGCAGTCGCATTAGACCGCTGCTTACTT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATAGTTAGATATGGAAAAACTTCCATACAGTCGTACCTCTGCAGTCGCATTAGACCGCTGCTTACTT 720

Qy   721 GTCTACTCCAAAAAGATCCCATGCCCAGTGGCAGAAAGAGGAGAAATGATAAAGAATTTTACC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTCTACTCCAAAAAGATCCCATGCCCAGTGGCAGAAAGAGGAGAAATGATAAAGAATTTTACC 780

Qy   781 TTGAATGAATTTTGGGCTTCATGATTCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TTGAATGAATTTTGGGCTTCATGATTCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCT 840

Qy   841 CTAGCTCGGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGTATAAAAATAGTATC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CTAGCTCGGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGTATAAAAATAGTATC 900

Qy   901 TATAGTTGGGCCTGGAAGCCTTTGGGGATTGGTTAAATTAGGAAGAAGACCACCTACTTTTGGAATGAGAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TATAGTTGGGCCTGGAAGCCTTTGGGGATTGGTTAAATTAGGAAGAAGACCACCTACTTTTGGAATGAGAT 960

Qy   961 GTGGAGAAGGCAATTTATGAAGGCTAGCTCTGAACTCTTCAGTCATCAGAAAACAAAAGGCATCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GTGGAGAAGGCAATTTATGAAGGCTAGCTCTGAACTCTTCAGTCATCAGAAAACAAAAGGCATCT 1020

Qy   1021 TTACTTGTATCAAAATCAAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACTTGTATCAAAATCAAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCA 1080

Qy   1081 TCTGTTCTAGCACAGTACTCACCTCACCTCAGCAATTAGCAGGAGAAGAAGAAATTGGAGTGTTTTCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTGTTCTAGCACAGTACTCACCTCACCTCAGCAATTAGCAGGAGAAGAAGAAATTGGAGTGTTTTCT 1140

Qy   1141 TATGGTTCTGGTTTGGCTGCCACTCGTACTCCTTAAAGTCACACAAGAGTACACCG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TATGGTTCTGGTTTGGCTGCCACTCGTACTCCTTAAAGTCACACAAGAGTACACCG 1200

Qy   1201 GGGTCTGCGTCTTGATAAAATAACAGAGAAGTTTATGTGATGATGTGAAAATCAAGGCTTGATTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGTCTGCGTCTTGATAAAATAACAGAGAAGTTTATGTGATGATGTGAAAATCAAGGCTTGATTCA 1260

Qy   1261 AGAACTGGTGTGGCACCAGATGTTCGCTGAAAAACATGAGAGCTTCAGAGGAGGACACCCAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAACTGGTGTGGCACCAGATGTTCGCTGAAAAACATGAGAGCTTCAGAGGAGGACACCCAT 1320

Qy   1321 CATTTGGTCAACTATATTCCCAGGGGTTCAATAGATTCACTCTTTGAACGGAACAGTGGTAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CATTTGGTCAACTATATTCCCAGGGGTTCAATAGATTCACTCTTTGAACGGAACAGTGGTAC 1380

Qy   1381 TTAGTTAGGGTGGATGAAGAAGGCACGAGAAGAACTTACGGCTCGGCGCGTCCCCACTCCAAAATGAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTAGTTAGGGTGGATGAAGAAGGCACGAGAAGAACTTACGGCTCGGCGCGTCCCCACTCCAAAATGAT 1440

Qy   1441 GACACTTTGGATGAAGGAGGTAGGACACTTGTTGCATTCAACTGCAACTGCAGCATAATTCCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GACACTTTGGATGAAGGAGGTAGGACACTTGTTGCATTCAACTGCAACTGCAGCATAATTCCA 1500

Qy   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCGCCACGCAGCAACCTGAAAGCTGTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCGCCACGCAGCAACCTGAAAGCTGTC 1560

Qy   1561 ATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCCAGCAGCAGCAGCTGGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCCAGCAGCAGCAGCTGGGC 1620

Qy   1621 ATGGGGTGGGGGTATGGGGAACAGTTGGGAGGAATGGGGATAATTTTTAAAGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGGGTGGGGGTATGGGGAACAGTTGGGAGGAATGGGGATAATTTTTAAAGGA 1680

Qy   1681 TTACAGTTAGTAAAATTTTTATGTGACTGACATGGAGCCTGGATGACTATCGTGTACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTACAGTTAGTAAAATTTTTATGTGACTGACATGGAGCCTGGATGACTATCGTGTACTT 1740

```
Db   1681 TTACATGTTATGTAAATTTTTATGTGACTGACATGGAGCCTGGATGACTAATCCTGTACTT 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1741 GGGAAAGTCTCTTTTGCTCTATTTGCTGACATGCTTCCGTTGTGGTCTGGCCAATGCCAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGGAAAGTCTCTTTGCTCTATTTGCTGACATGCTTCCGTTGTGGTCTGGCCAATGCCAA 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1801 ATGTACTCGGAATCGATGTTAAGGGCTCTGTAAAACTTCATACCTTTCGGCCATTTGTATG 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATGTACTCGGAATCGATGTTAAGGGCTCTGTAAAACTTCATACCCTTTCGGCCATTTGTATG 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1861 CATGATGTTTGGTTTTAAACATGGTATATAATGAATTGTGTACTTCGTGTCAGAAGAAAGCA 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CATGATGTTTGGTTTTAAACATGGTATATAATGAATTGTGTACTTCGTGTCAGAAGAAAGCA 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1921 GAGGTACTAATCCTCCAATTAAAAAATTTTTTTAACATGTAAAAAAAAAAAAAAAAAAA 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGGTACTAATCCTCCAATTAAAAAATTTTTTTAACATGTAAAAAAAAAAAAAAAAAAA 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1981 AAAAAAAAAAAAAAAAAAAAAA 2002
          |||||||||||||||||||||
Db   1981 AAAAAAAAAAAAAAAAAAAAAA 2002
```

```
RESULT 2

ID   ABX93299 standard; cDNA; 2002 BP.
XX
AC   ABX93299;
XX
DT   05-JUN-2003 (first entry)
XX
DE   cDNA encoding human HMG-CoA synthase-like enzyme.
XX
KW   Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KW   HMG-CoA synthase; immune response; drug screening assay;
KW   pharmacogenomic analysis; chromosome 5; gene; 88.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .144
FT                   /*tag= a
FT   CDS             145. .1581
FT                   /*tag= b
FT                   /product= "HMG-CoA synthase-like enzyme"
FT   3'UTR           1582. .2002
FT                   /*tag= c
XX
PN   US2002173018-A1.
XX
PD   21-NOV-2002.
XX
DR   WPI; 2003-352594/33.
DR   P-PSDB; ABU08379.
XX
PF   12-JUL-2002; 2002US-00193295.
XX
PR   29-MAR-2001; 2001US-0081993.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Gong F,  Yan C,  Di Francesco V,  Beasley EM;
XX
PS   Claim 22; Fig 1A; 65pp; English.
XX
PT   New isolated human synthase peptide and gene encoding the enzyme, useful
PT   as models for developing human therapeutic targets, aid in the
PT   identification of therapeutic proteins and as therapeutic targets.
XX
CC   The present invention relates to the isolation of a novel human enzyme
CC   that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
CC   methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also
```

```
CC   discloses polynucleotide sequences encoding the novel enzyme of the
CC   invention. Both the polypeptide and polynucleotide sequences are useful
CC   as models for the development of human therapeutics, for identifying
CC   therapeutic proteins, as targets for development of human therapeutic
CC   agents, and as query sequences to perform a search against sequence data
CC   bases to identify other family members of related sequences. The
CC   polypeptide is useful in assays designed to raise antibodies or to elicit another immune
CC   response, as a reagent in assays designed to quantitatively determine
CC   levels of the protein in biological fluids, as markers for tissues in
CC   which the corresponding protein is preferentially expressed, in drug
CC   screening assays, in cell-based or cell-free systems to identify
CC   compounds that modulate synthase activity of the protein in its natural
CC   state, or an altered form that causes the specific disease or pathology
CC   associated with the synthase, to screen a compound for the ability to
CC   stimulate or inhibit interaction between the synthase protein and a
CC   molecule that normally interacts with the synthase protein, and in
CC   diagnostic assays for qualitative changes in a human synthase nucleic
CC   acid that leads to a pathology, for testing an individual for a genotype
CC   that while not necessarily causing a disease, nevertheless affects the
CC   effectiveness of modulating compounds on the expression or activity of
CC   pharmacogenomic analysis. The polynucleotide is useful for monitoring the
CC   effectiveness of modulating compounds on the expression or activity of
CC   the human synthase gene in clinical trials or in a treatment regimen, in
CC   gene expression in cell, tissues and organisms. The present sequence
CC   treatment modality, and as antisense constructs to control human synthase
CC   encodes human HMG-CoA synthase-like enzyme. The gene encoding the enzyme
CC   is located on chromosome 5
XX
SQ   Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;
```

```
Query Match             100.0%;  Score 2002;  DB 7;  Length 2002;
Best Local Similarity   100.0%;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CGCCTCCCACGCGACTCTCGGCCAGTGCCCGGAGAGTGCGGTGGGGTGGGCCCGTATAAAAGCTGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCCTCCCACGCGACTCTCGGCCAGTGCCCGGAGAGTGCGGTGGGGTGGGCCCGTATAAAAGCTGGT 60

Qy     61 AGCGAAGGGGAGGGGCCCGCGGGACTGTCCTTTCGTTGGTCGCTCACTCCCTTTCCTCTGCTGCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGCGAAGGGGAGGGGCCCGCGGGACTGTCCTTTCGTTGGTCGCTCACTCCCTTTCCTCTGCTGCCG 120

Qy    121 CTCCGGTCACCGCTTGCTCTTCTTCACCATGGCTCGGAGTCGGGGTTGGCGGCTATAAAGCTGGT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCGGTCACCGCTTGCTCTTCTTCACCATGGCTCGGAGTCGGGGTTGGCGGCTATAAAGCTGGT 180

Qy    181 TGGCCAAAAGATGTGGGAATTGTTGTTCCCCCTTGAGACTTATTTTCCTTCCAATATGTTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGCCAAAAGATGTGGGAATTGTTCCCCCTTGAGACTCTATTTTCCTTCTCAATATGTTGAT 240

Qy    241 CAAGCAGAGAGTTGGAAAAAATATGATGGTGTRAGATGGTGGAAAAGTATACCATTGGCTTGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAGCAGAGAGTTGGAAAAAATATGATGGTGTRAGATGGTGGAAAAGTATACCATTGGCTTGGGC 300

Qy    301 CAGGCCAAGAGAGATGGGGCTTCTGCACAGATAGGAGAAGATATTAACTCTCTTTGCATGACTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGCCAAGATGGGGCTTCTCGCACAGATAGGAGAAGATATTAACTCTCTTTGCATGACTGTG 360

Qy    361 GTTCAGAATCTTATGGAGAGAAAATAAACCTTTCCTATGATTGCATTGGGCGCGGCTGGAAGTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTCAGAATCTTATGGAGAGAAAATAAACCTTTCCTATGATTGCATTGGGCGCGGCTGGAAGTT 420

Qy    421 GGAACAGAGACAATCATCGACAAATCAAAGTCTGTGAAGACTAATTGATTCCAGCTGGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGAACAGAGACAATCATCGACAAATCAAAGTCTGTGAAGACTAATTGATTCCAGCTGTTT 480

Qy    481 GAAGAGTCTGGGAATACAGATTATAGAAGGAATCGACAACTAATGCATGCTATGGGAGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAAGAGTCTGGGAATACAGATTATAGAAGGAATCGACAACTAATGCATGCTATGGGAGGC 540

Qy    541 ACAGCTGCGCTGTCTTCAAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGGATGGGC CGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACAGCTGCGCTGTCTTCAAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCCGT 600
```

```
Qy    601 GGGACACATATGCAAACATGCCTATGATTTTTACAAGCCTGATATGCTTATCTGAATATCCT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGGACACATATGCAAACATGCCTATGATTTTTACAAGCCTGATATGCTTATCTGAATATCCT 660

Qy    661 ATAGTAGATTGGAGAAACTCTCATACAGTGTCACTCAGTGATGATCGCTGCTACTCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATAGTAGATTGGAGAAACTCTCATACAGTGTCACTCAGTGATGATCGCTGCTACTCT 720

Qy    721 GTCTACTCTGCAAAAAGATCCATGCCCACGTGGCAGAAAATGATAAACGATTTTACC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTCTACTGCGAAAAAAGATCCATGCCCAGTGGGCAGAAAATGATAAAAGATTTTACC 780

Qy    781 TTGAATGATTTTGGCTTCATCATTCTTCACTCACCATATTGTAAACTGGTTCAGAAATCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCT 840

Qy    841 CTAGCTCCGATGTTGCTGAAATGACTTCCTTAATCACCAGAATAGAGATAAAATAGTTATC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTAGCTCCGATGTTGCTGAAATGACTTCCTTAATCACCAGAATAGAGATAAAATAGTTATC 900

Qy    901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATACAGAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATACAGAT 960

Qy    961 GTGGAGAAGGCATTTATGAAGGCTAGCTCTCGAACTCTTCAGTCTTCAGTCTTCCCTTGCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTGGAGAAGGCATTTATGAAGGCTAGCTCTCGAACTCTTCAGTCTTCAGTCTTCCCTTGCA 1020

Qy   1021 TTACTTGTATCAAATCAAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACTTGTATCAAATCAAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080

Qy   1081 TCTGTTCTTAGCACAGTACTCACCTCCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTGTTCTTAGCACAGTACTCACCTCCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTCT 1140

Qy   1141 TATGGTTCTGGTTTGGCTGCCACTCGTGTACTCTGTGTACTCTTAAAGTCACACAAGATGG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TATGGTTCTGGTTTGGCTGCCACTCGTGTACTCTGTGTACTCTTAAAGTCACACAAGATGG 1200

Qy   1201 GGGTCTGCTCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTTAAATCAAGGCTTGATTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGTCTGCTCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTTAAATCAAGGCTTGATTCA 1260

Qy   1261 AGAACTGGGTGTGGCCACCAGATGTCTTCGGCTGAAAAACATGAAGCTTCAGAGAGGACACCCAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAACTGGGTGTGGCCACCAGATGTCTTCCGCTGAAAAACATGAAGCTTCAGAGAGGACACCCAT 1320

Qy   1321 CATTTGGTCAACTATATTATTCCCAGGGTTCAAATAGATTCACTCTTTGAAGACAGTGGTAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CATTTGGTCAACTATATTATTCCCAGGGTTCAAATAGATTCACTCTTTGAAGACAGTGGTAC 1380

Qy   1381 TTAGTTAGGGTGGATGAAAAAGCACAGAAGAACTTACGCTCGGCGTCTCCCACTCCAAATGAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTAGTTAGGGTGGATGAAAAAGCACAGAAGAACTTACGCTCGGCGTCTCCCACTCCAAATGAT 1440

Qy   1441 GACACTTTGGATGAAGGAGGTAGGACTTGTGTCATTCAAACTAGCAACTGAGCATATTCCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GACACTTTGGATGAAGGAGGTAGGACTTGTGTCATTCAAACTAGCAACTGAGCATATTCCA 1500

Qy   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGAACCTGAAGCAGCTGTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGAACCTGAAGCAGCTGTC 1560

Qy   1561 ATTAGTGAATGGGAACATTAAGATACTCTGTGAGGTTGCAAGACTTCAGGGTGGGGTGGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATTAGTGAATGGGAACATTAAGATACTCTGTGAGGTTGCAAGACTTCAGGGTGGGGTGGGC 1620

Qy   1621 ATGGGGTGGGGGTATGGGGAACAGTTGGAGGAATGGGATAATGTGGGATAATTTTAAAGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGGGTGGGGGTATGGGGAACAGTTGGAGGAATGGGATAATGTGGGATAATTTTAAAGGA 1680

Qy   1681 TTACATGTTATGTAAATTTTATGTGACTGAACTGAACATGGGAGGCCTGGATGGACTATCGTGTACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTACATGTTATGTAAATTTTATGTGACTGAACTGAACATGGGAGGCCTGGATGGACTATCGTGTACTT 1740

Qy   1741 GGGGAAGAGTCTTGCTCTATTTGCTGACATGCTCCTGTTGTTGGTCTGGCCAATGCCAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGGGAAGAGTCTTGCTCTATTTGCTGACATGCTCCTGTTGTTGGTCTGGCCAATGCCAA 1800

Qy   1801 ATGTACTCGAATGATGTTAAGGGCTCTTGTAAAACTTCATACCTCTTTGGCCATTTGTATG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATGTACTCGAATGATGTTAAGGGCTCTTGTAAAACTTCATCACCTCTTTGGCCATTTGTATG 1860

Qy   1921 GAGGTACTAATCTCCAATTAAAAAAATTTTTAACATGTAAAAAAAAAAAAAAAAAAGCA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGGTACTAATCTCCAATTAAAAAAATTTTTAACATGTAAAAAAAAAAAAAAAAAAA 1980

Qy   1981 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2002
          ||||||||||||||||||||||||||
Db   1981 AAAAAAAAAAAAAAAAAAAAAAAAA 2002
```

RESULT 3

ADE76933

ID   ADE76933 standard; cDNA; 3722 BP.
XX
AC   ADE76933;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human cDNA differentially expressed in a liver disorder #70.
XX
KW   human; ss; gene; liver disorder; hyperlipidaemia; hypertension;
KW   type II diabetes; tumour; liver; inflammatory disorder;
KW   immune response disorder; high-throughput screening;
KW   differential gene expression; gene therapy.
XX
OS   Homo sapiens.
XX
PN   US2003108871-A1.
XX
PD   12-JUN-2003.
XX
PF   30-JUL-2001; 2001US-00919039.
XX
PR   28-JUL-2000; 2000US-0222113P.
XX
PA   (KASE/) KASER M R.
XX
PI   Kaser MR;
XX
DR   WPI; 2004-031227/03.
XX
PS   Claim 1; SEQ ID NO 98; 41pp; English.
XX
PT   Composition comprising several cDNAs that are differentially expressed in
PT   treated human C3A liver cell cultures, useful for treating liver
PT   disorders.
CC   The invention relates to a composition comprising several cDNAs that are
CC   differentially expressed in a liver disorder. The composition is useful
CC   for treating liver disorder such as hyperlipidaemia, hypertension, type
CC   II diabetes, tumours of the liver and disorders of the inflammatory and
CC   immune response. The composition is useful for a high-throughput method
CC   of screening several molecules or compounds to identify a ligand which
CC   specifically binds a cDNA. A protein encoded by the cDNA is useful for a
CC   high-throughput method for using a protein to screen several molecules or
CC   compounds to identify at least one ligand which specifically binds the
CC   protein which involves combining the protein encoded by the cDNA with
CC   several of molecules or compounds under conditions to allow specific
CC   binding, and detecting specific binding between the protein and a

```
CC   molecule or compound, therefore identifying a ligand which specifically
CC   binds the protein. The composition is useful for detecting and
CC   quantifying differential gene expression, can be used in gene therapy, to
CC   formulate prognosis and to design a treatment regimen and to monitor the
CC   efficacy of treatment. The present sequence represents a cDNA
CC   differentially expressed in a liver disorder.
XX
SQ   Sequence 3722 BP; 1121 A; 645 C; 860 G; 1094 T; 0 U; 2 Other;

Query Match             84.6%;  Score 1693;  DB 10;  Length 3722;
Best Local Similarity   92.7%;  Pred No. 0;
Matches 1841;  Conservative    0;  Mismatches   20;  Indels 126;  Gaps    1;

Qy        134 GCTCTTTCACCATGCCCTGGATCACTCACTTCCTTTTGAATGCAGAAGCTTGCTCGGCCAAAAGAATG 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 GCTCTTTCACCATGCCCTGGATCACTCACTTCCTTTTGAATGCAGAAGCTTGCTCGGCCAAAAGAATG 297

Qy        194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCTCAAATATGTTGATCAAGCAGAGTTGG 253
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        298 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCTCAAATATGTTGATCAAGCAGAGTTGG 357

Qy        254 AAAAATATGATGGTGTAGAATGCTGGAAATATTACCATTGGCTTGGCCCAAGATGG 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 AAAAATATGATGGTGTAGAATGCTGGAAAGTATTACCATTGGCTTGGCCCAAGATGG 417

Qy        314 GCTTCTGCACAGATAGAGGAAGAATATTAACTCTCTTTGCATGAGAATCTTA 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 GCTTCTGCACAGATAGAGGAAGAATATTAACTCTCTTTGCATGAGAATCTTA 477

Qy        374 TGGGAGAAAATAAACCTTTCCTATGGATTGCATTGGGGCTGGGAACAGAGACAA 433
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 TGGGAGAAAATAAACCTTTCCTATGGATTGCATTGGGGCTGGGAACAGAGACAA 537

Qy        434 TCATCGACAAAATCAAAAGTCTGTGAAGACTAATTTTGATGCAGCGTGTTTGAAGAGTCTGGGA 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        538 TCATCGACAAAATCAAAAGTCTGTGAAGACTAATTTTGATGCAGCGTGTTTGAAGAGTCTGGGA 597

Qy        494 ATACAGATAGAAGGAAATCGACAACAACTAAATGCATTGGAGAGGCAGCAGCTGCTGTCT 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        598 ATACAGATAGAGAGGAATCGACAACAACTAAATGCATTGGAGAGGCAGCAGCTGCTGTCT 657

Qy        554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT 591
              ||||||||||||||||||||||||||||||||||
Db        658 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATA 717

Qy        592 ------------------- 591
              
Db        718 CAGGAGATATTGCTGTATATGCCACAGGTGGAGTTGGACGAG 777

Qy        592 ------------------- 591
              |||||||||||||||
Db        778 TACCTCTGCTAATTGGGGCAAAATGCTCCTTAATTTTGAAACGAGGGCTTCGTGGGACAC 837

Qy        608 ATATGCAACATGCCTATGATTTTTACAAGCCTATGTAG 667
              ||||||||||||||||||||||||||||||||||||
Db        838 ATATGCAACATGCCTATGATTTTTACAAGCCTATGTAG 897

Qy        668 ATGGAAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCCGCTACTT 727
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        898 ATGGAAAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCCGCTACTT 957

Qy        728 GCAAAAGAGTCCATGCCCAGTGGGCAGANTGAAAGAGGGAAATTTACCTTGAATG 787
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        958 GCAAAAGAGTCCATGCCCAGTGGGCAGAAATGATAAAGATTTTACCTTGAATG 1017

Qy        788 ATTTTGGCTTCATGATCTTTCACTCACCACATATTGTAAACTGGGTTCAGAAATCTCTAGCTC 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1018 ATTTTGGCTTCATGATCTTTCACTCACCAATATTGTAAACTGGGTTCAGAAATCTCTAGCTC 1077

Qy        848 GGATGTTGCTGAATGAGCTTCCTTAATGACCAGAGAATAGAGGATAAAAATAGTATATCTATAGTG 907
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1078 GGATGTTGCTGAATGACTTCCTTAATGACCAGAGAATAGAGGATAAAAATAGTATATCTATAGTG 1137

Qy       1018 ATTTTGGGCTTCATCGATCTTTCACACCAATATTGTAAACTGGGTTCAGAAATCTCTAGCTC 1077

Qy        968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCGTCAGAAAACAAAGGCATCTTTACTTG 1027
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 GCCTCGGAAGCCTTTGGGGGATGTTAAATTAGAAGACACCTACTTTGATGTGGAGA 1197

Qy       1088 TAGCACAGTACTCACCTCAGCCAATTAGCACGGGGAACAGAGAATTGGACGTGTTGTTTTTCTTANGGTT 1147
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 TATCAAAATCAAAAGTGGAAAATGTGTACACATCTTCAGTATATGGGTTCCCTTGCATCGTGTTC 1317

Qy       1198 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCGTCAGAAAACAAAGGCATCTTTACTTG 1257

Db       1318 TAGCACAGTACTCACCTCAGCCAATTAGCACGGGGAACAGAGAATTGGACGTGTTGTTTTTCTTANGGTT 1147

Qy        908 GCCTCGGAAGCCTTTGGGGGATGTTAAATTAGAAGACACCTACTTTGATGTGGAGA 967
```

(The remainder of the page continues the sequence alignment in two columns of paired Qy / Db rows down to)

```
Qy       1928 TAATCTCCAATTAAAAAATTTTTTTAACATGTGAAGAATTTGTACTTTGAACAAAGAGATT 1987
Db       2158 TAATCTCCAATTAAAAAATTTTTTTAACATGTGAAGAATTTGTACTTTGAACAAAGAGATT 2217

Qy       1988 AAAAAAAA 1994
```

Db   2218 ACAGAAA 2224
            | | |||

RESULT 4
ID   AAS84743 standard; cDNA; 2540 BP.
XX
AC   AAS84743;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20547.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO20017SO67-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 20547; 103pp; English.
XX

The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
activity of (II) or to treat disease states involving (II). (II) is
useful for generating antibodies against it, detecting or quantitating a
polypeptide in tissue, as molecular weight markers and as a food
supplement. (II) and its binding partners are useful in medical imaging
of sites expressing (III). (I) and (II) are useful for treating disorders
involving aberrant protein expression or biological activity. The
polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
coding sequences of the invention. Note: The sequence data for this
patent did not appear in the printed specification, but was obtained in
electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

SQ   Sequence 2540 BP; 732 A; 474 C; 621 G; 713 T; 0 U; 0 Other;

Query Match   80.2%; Score 1605.2; DB 5; Length 2540;
Best Local Similarity 89.6%; Pred. No. 3.3e-305;
Matches 1921; Conservative 0; Mismatches 18; Indels 205; Gaps 10;

Qy   25 GCCGGAGTCGGGTGGGTTGGCGGCCTATAAAGCTGGTTAGTCGAAGGGAGGCGCGCCCGGACT 84
          ||||||||||||||||||||||||||||||||||||| |||||| | |||||||| |||||||| ||
Db   5 GTCGGAGTCGGGTGGGTTGGCGGCCTATAAAGCTGGTTGGGCGAAGGGAGGGGACGCGCCGCGGACT 64

Qy   85 GTCCTTTCGTGGCTCACTCCCTTTCCTCTGCTGCTGCCGCTGCCGCTTCGGTCACGGCTT------------- 133
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   65 GTCCTTTCGTGGCTCACTCCCTTTCCCTCTGCTGCTGCCGCTGCCGCTTCGGTCACGGCTT----------GCTCTTTCACCA 145

Qy   134 -- 133
Db   125 AGGAAACAGTGACAGACCTGGAGACTCCAGTTCTCTATCCTTCACACACAGTCTTTCACCA 184

Qy   146 TGCCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGGCCAAAAAGAGATGTGGGGAATTGTTG 205
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 TGCCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGGCCAAAAAGAGATGTGGGGAATTGTTG 244

Qy   206 CCCTTGAGATCTTATTTTCCTTTCTCAAATATGTTGATCAAGCAGAGCAGAGTTGGGAAAATATGATG 265
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 CCCTTGAGATCTTATTTTCCTTTCTCAAATATGTTGATCAAGCAGAGCAGAGTTGGGAAAATATGATG 304

Qy   266 GTGTAGATGTGCTTGGAAAGTAТACCATTGGCTTGGGCGCCAGGGCCCAAGAGATGGGCTTCGCACAG 325
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   305 GTGTAGATGCGTGGAAAAGATAССАТТGGCTTGGGCCAGGGCCAACGATGGGCTTCGCACACA 364

Qy   326 ATAGAGAAGAGTAATAAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGGAGAGAAATA 385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 ATAGAGAAGAGTAATAAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGGAGAGAAATA 424

Qy   386 ACCTTTCCTATGATTGCATTGGGCGGCCTGGGAAGGTTGGAACAGAACAATCATCGACAAAAT 445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425 ACCTTTCCTATGATTGCATTGGGCGGCGCTGGGAAGGTTGGAACGAGACAATCATCGACAAAT 484

Qy   446 CAAAGTCTGTGAAGAGTAATTGAATGCAGTGTTTGAAGAGTCTGGGAATACAGATATAG 505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   485 CAAAGTCTGTGAAGAGACTAATTTGATGCAGTGTTTGAAGAGTCTGGGAATACAGATATAG 544

Qy   506 AAGGAATCGACAACTAATGCTATGCAGCTGCTTGCTGTCTTCTTCAATGCTGTTA 565
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   545 AAGGAATCGACAACTAATGCTATGGAGGCACAGCTGCTGCTGTCTTCTTCAATGCTGTTA 604

Qy   596 ---TTCGTGGGGACACA 608
Db   605 ACTGGATTGAGTCCAGCTCTTGGGATGGCACGGTATGCCCTGGGGTAGTGTGCAGGAGATATT 664

Qy   566 ACTGGATTGAGTCCAGCTCTTGGGATGGGGC---------------------- 595
          ||||||||||||||||||||||||||||||
Db   ...

Qy   605 ACTGGATTGAGTCCAGCTCTTGGGATGGGGC---------------------- 595

Qy   485 CAAAGTCTGTGAAGAGACTAATTTGATGCAGTGTTTGAAGAGTCTGGGAATACAGATATAG 544

Qy   596 -- 595
Db   665 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGGAGTTGAGCAGTAGCTCTTGC 724

Qy   665 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGGAGTTGAGCAGTAGCTCTTGC 724

Qy   725 TAAAATTGGGCGGCCAAATAGCTCCTTTTAATTTTTTGAAACCGAGGGGGTTTCGTGGGCACA 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   785 TATGCAACATGCCCTATGAGTTTTTACAAGCCCTGATAATGCTATCTGAATAATCCTATAGTAGA 844

Qy   609 TATGCAACATGCCCTATGAGTTTTTACAAGCCCTGATAATGCTATCTGAATAATCCTATAGTAGA 668

Qy   729 CAAAAGAATCCATCGCCAGTGGCAGAAGAGGGAAATGATAGAGATTTTTACCTTGAATGA 788
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   905 CAAAAAGAATCCATCGCCCAGTGGCAGAAGAGGGAAATGATAGAGATTTTTACCTTGAATGA 964

Qy   845 TGGGAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAACCGCGCTAGTTCCTGCTCTACTG 904

Qy   669 TGGAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAACCGCGCTAGTTCGTCTCTACTG 728

Qy   789 TTTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCG 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   965 TTTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCG 1024

Qy   849 GATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAAAAATAGTATCTATAGTGG 908

Qy   965 GATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAAAAATAGTATCTATAGTGG 1024

Qy   909 CCTGGAAGCCTTTGGGAGATGTTAAATTAGAAGACACCTACTTACTTTGATAGAGATGGAGAA 968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1025 GATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAAAAATAGTATCTATAGTGG 1084

Qy 1025 GATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAAAAATAGTATCTATAGTGG 1084

Db 1085 CCTGGAAGCCTTTGGGAGATGTTAAATTAGAAGACACCTACTTACTTTGATAGAGATGGAGAA 1144

```
RESULT 5
ACC62336
ID   ACC62336 standard; cDNA; 1650 BP.
XX

Qy    969 GGCATTTATGAAGGCTAGCTCTGAAACTCTTCAGTCAGAGAAACAAAAGGCATCTTTACTTGT 1028
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1145 GGCATTTATGAAGGCTAGCTCTGAAACTCTTCAGTCAGAGAAACAAAAGGCATCTTTACTTGT 1204

Qy   1029 ATCAAAATCAAAATATGGAAATATGTACACATCTTCAGTATATATGGTTCCCTTGCATCTGTTCT 1088
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1205 ATCAAAATCAAAATATGGAAATATGTACACATCTTCAGTATATATGGTTCCCTTGCATCTGTTCT 1264

Qy   1089 AGCACAGTACTCACCTCAGCAATTAGCACGGGAAGAGAATTGGGAGTGTGTTTTCTTATGGTTC 1148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1265 AGCACAGTACTCACCTCAGCAATTAGCACGGGAAGAGAATTGGGAGTGTGTTTTCTTATGGTTC 1324

Qy   1149 T-GGTTTGGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTCACCCGGGGTCTG 1207
          | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1325 TGGGTTTGGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTCACCCGGGGTCTG 1384

Qy   1208 CTCTTGAT-AAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACT 1266
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1385 CTCTTGATAAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACT 1444

Qy   1267 GGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTTCAGAGAGAGGACACCCATCATTTG 1326
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1445 GGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTTCAGAGAGAGGACACCCATCATTTG 1504

Qy   1327 GTCAACTATATTCCCCAGGGTTCAATAGAGATTCACTCTTTTGAAGGAACCTGGTACTTAGTT 1386
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1505 GTCAACTATATTCCCCAGGGTTCAATAGAGATTCACTCTTTTGAAGGAACCTGGTACTTAGTT 1564

Qy   1387 AGGGTGGATGAAAAGCACAGAGAGAAC-TTACGCTGGCTTCGGGCTCCCACTCCAAATGATGACAC 1445
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   1565 AGGGTGGATGAAAAGCACAGAGAGAACTTACGCTGGCTTCGGGCTCCCACTCCAAATGATGACAC 1624

Qy   1446 TTTGGATGAAGGAGTAGGACTTGTTGCATTCAAACATAGCAACTGAGCANTATTCCAAGCCC 1505
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   1625 TTTGGATGAAGGAGTAGGACTTGTTGCATTCAAACATAGCAACTGAGCATTATTCCAAGCCC 1684

Qy   1506 TGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAAACCTGAAAGCAGCTGTCATTAG 1565
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1685 TGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAAACTGAGCAGCAGCTGTCATTAG 1744

Qy   1566 TAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGAACTTCAGGGGTGGGGTGGGCCATGGG 1625
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1745 TAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGAACTTCAGGGGTGGGGTGGGCCATGGG 1804

Qy   1626 GTGGGGGTATGGGAACAGTTGGAGGAATGGGATTATTCTGGGGATAATTTTAAAGGATTACA 1685
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1805 GTGGGGGTATGGGAACAGTTGGAGGAATATCTGGGGATAATTTTAAAGGATTACA 1864

Qy   1686 TGTTATGTGAAATTTTAATGTGAGTGACACAT-GGAGCCTCGGATGACTATATCGTTGTACTTGGGA 1744
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   1865 TGTTATGTAAATTTTTATGTGACATGGGACCCTGGATGACTATCGTGTACTTGGGA 1924

Qy   1745 AAGTCTCTTTGCTCTATTTGCTGACATG--CTTCCGTGTTGTGGGTCGTTGGGCCAATGCC--AAA 1801
          |||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||  |||
Db   1925 AAGTCTCTTTGCTCTATTTGCTGACATGGCCTTCCGTGTTGTGGGTCGTTGGGCCAAATGCCCAAAT 1984

Qy   1802 TGTACTCGAATGAGTGTTAAGGGCTTCGTGTAAAACTTCATACCTCTTTGGGCCCATTTGTATGC 1861
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1985 GTACCTCGAATGAGATGTTAAGGGCTTCGTGTAAAACTTCATACCTCTTTGGGCCCATTTGTATGC 2044

Qy   1862 ATGATGTTTGGGTTTTTTAAACATGTGTATAAT-GAATTGTGTTACTTCTGTCAGAA-GAAAGC 1919
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||
Db   2045 ATGATGTTTGGGTTTTTTAAACATGGTATAATGGAATTGTGTTACTTCTGTCGAAGGAAAGC 2104

Qy   1920 AGAGGTACTAAATCTCCAATTAAAAAAAATTTTTAACATGTAAAAA 1963
          ||||||||||||||||||||||||||     |||||||||||||||
Db   2105 AGAGGTACTAAATCTCCAATTAAAAAATTTTTTAACATGTAAGAA 2148
```

XX

CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the NOVX protein in human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic disorders.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein in comparison with
CC   the human NOV35b protein in the exemplification of the present invention

XX

SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match   68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity   92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative   0; Mismatches   5; Indels 126; Gaps   1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGTTTGAATGCAGAAGGCTTGCTTGCCAAAGATG 193
          |||||||||||||||||||||||||||||||| ||||||||||||| |||||| ||||||||||
Db   11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAAATGCAGAAGCTTGCTCTGGCCCAAAAGATG 70

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCTCAAATATGTTGATCAAGCAGAGTTGG 253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 TTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAAATATGATGGTTCTAGAGTGCTTGGGAAAGTATACCATTGGCTTGGCCTTGGGCCCAGGCCCAAGATGG 313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAAATATGATGGTTCTAGAGTGCTTGGGAAAGTATACCATTGGCTTGGCCTTGGGCCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAAGAAGATATTAAACTCTCTTTGCCATCGTGTTCAGAATCTTA 373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAAGAATATTAAACTCTCTTTGCCATGAATGACTCTGGTTCAGAATCTTA 250

Qy   374 TGGAGAGAAAATAAACCTTTCCTATGATTGCCATTGGGCGGCCGCTGGAACAGAGACAA 433
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGAGAAAATAAACCTTTCCTATGATTGCCATTGGGCGGCCGCTGGAACAGAGACAA 310

Qy   434 TCATCGACAAANTCAAAAGTCTGACACACAACTAATGCAATGCTACGTGCTGCTCTCT 493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAAGTCTGACACACAACTAATTGATGCAGCGTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAATCGACACACAACTAATGCATGCTATGGAGGCACAGCTGCTCTGT 553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 ATACAGATATAGAAGGAATCGACACAACTAATGCTATGGAGGCACAGCTGCTCTGTCT 430

Qy   554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------------- 591
          ||||||||||||||||||||||||||||||||||||||
Db   431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490

Qy   592 -- 591
Db   491 CAGGAGATATTGCTCTATATGCCACAGGAAATGCTAGAACCTAACTAGGCTGGAGTTGGAGCAG 550

Qy   592 --GGGCTTCGTGGGACAC 607
                                                          ||||||||||||||||
Db   551 TAGCTCTTGCCTCAATTGGGCCAAATGCTCCTTTAATTTTGAACAGAGGGCTTCGTGGGACAC 610

Qy   608 ATATGCAACATGCCTATGATTTTACAAGGCCTGATATCTGAATATCCTATAGTAG 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   608 ATATGCAACATGCCTATGATTTTACAAGGCCTGATATCTGAATATCCTATAGTAG 670

RESULT 6

```
ID   ACC62328
XX
AC   ACC62328 standard; cDNA; 1650 BP.
XX
AC   ACC62328;
XX
DE   Human NOV41c encoding cDNA SEQ ID NO:185.
XX
DT   23-JUN-2003 (first entry)
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VV, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ABR54259.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 241; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins, given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match             68.4%;  Score 1370;  DB 7;  Length 1650;
Best Local Similarity   92.0%;
Matches 1509; Conservative   0; Mismatches   5;  Indels 126;  Gaps   1;

Qy        134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATCAGAGAAGCTTCCTGGCCAAAAGATG 193
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATCAGAGAAGCTTCCTGGCCAAAAGATG 70

Qy        194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTTCCAATATGTTGATCAAGCAGAGTTGG 253
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         71 TTGGGATTGTTGCCCTTGAGATCTATTTTCCTTCTCCAATAATGTTGATCAACCAGAGTTGG 130

Qy        254 AAAAATATGATTGGTGGTAGATCTGGAAAGTATATACCATTGGCTTGGCCAGGCCAAGATGG 313
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 AAAAAATATGATGGGTGGTAGATCTGGAAAGTATACCATTGGGTGGGCCAGGCCAAGATGG 190

Qy        314 GCTTCTGCACAGATAGAGAGAGATATTAACTCTCTTCTTTGCATCACTGGTGGTTCAGAATCTTA 373
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        191 GCTTCTGCACAGATAGAGAGAGATATTAACTCTCTTCTTTGCATGACTGGTGGTTCAGAATCTTA 250

Qy        374 TGGGAGAAAATAAACTTTCCTATAGATTGCATTGGGCCGGCGTGGGAAGTTGGAACAGAGACA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        251 TGGGAGAAAATAAACCTTTCCTATAGATTGCATTGGGCGGGCTTGGAAGTTGGAACAGACAA 310

Qy        434 TCATCGACAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 493
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        311 TCATCGACAAAATCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 370

Qy        494 ATACAGATATAGAAGGAAATCGACACAACTAATGCATGGAGGACACACTGTGTCT 553
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        371 ATACAGATATAGAAGGAAATCGACACAACTAATGCATGGAGGCACACAGCTGTGTCT 430

Qy        554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTCTGGGAT--------------- 591
              |||||||||||||||||||||||||||||||||||||||
Db        431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTCTGGGATGGGACGGTATGCCCTGGTAGTTG 490

Qy        592 ---------------------------------------------------- 591
                                                                   |||||||||||||||
Db        491 CAGGAGATATTGCTGTATATGCCCAGGAAAATGCTAGACCTCACGGTTGGAGTTGGAGCAG 550

Qy        592 ---------------------------------GGGCTTCGTGGGACAC 607
                                               |||||||||||||||
Db        551 ...
```

```
Db   551  TAGCTGCTGCTAATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACAC  610
Qy   608  ATATGCCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAAATATCCTATAGTAG  667
Db   611  ATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAAATATCCTATAGTAG  670
Qy   668  ATGGAAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCGCTGCTATTCTGTCTACT  727
Db   671  ATGGGAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCGCTGCTATTCTGTCTACT  730
Db   728  GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATG  787
Db   731  GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATG  790
Qy   788  ATTTTGCTTCATGATCTTTCACTCACTCACCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC  847
Db   791  ATTTTGGCTTCATGATCTTTCATCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC  850
Qy   848  GGATGTTGCTGAATGACTTCCCTTAATGACCAGAATAGAGATAAAAATAGTATATCTATAGTG  907
Db   851  GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATATCTATAGTG  910
Qy   908  GCCTGGAAGCCTTTGGGGATGGTTAAATTAGAAGAAGCAACTTACTTTGAAGATGTGGAGA  967
Db   911  GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGAACACCTTACTTTGAAGATGTGGAGA  970
Qy   968  AGGGCATTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTG  1027
Db   971  AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTG  1030
Qy  1028  TATCAAATCAAAATGGAAAATATGTACACAATCTTCAGTATATGGTTCCCTTGCAATCTGTTC  1087
Db  1031  TATCAAATCAAAATGGAAAATATGTACACAATCTTCAGTATATGGTTCCCTTGCAATCTGTTC  1090
Qy  1088  TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTT  1147
Db  1091  TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTT  1150
Qy  1148  CTGGTTTGGCTGCCACACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG  1207
Db  1151  CTGGTTTGGCTGCCACACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG  1210
Qy  1208  CTCTTGATAAAATAACAGCAAGTTTATGTATCTTAAATCAAGGCTTGATTCAAGAACTG  1267
Db  1211  CTCTTGATAAAATAACAGCAAGTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG  1270
Qy  1268  GTGTGGCACCAGATGTCTTCGCTGCTGAAAACATGAAGCTCAGAGAGGACACCATCATTTGG  1327
Db  1271  GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCATCATTTGG  1330
Qy  1328  TCAACTATATTCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGTGGGTACTTAGTTA  1387
Db  1331  TCAACTATATTCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGTGGGTACTTAGTTA  1390
Qy  1388  GGGTGGATGAAAAAGCCACAGAGAACAGAACTTACGCTCGGCCGTTCCCACTCCAAATGATGACACTT  1447
Db  1391  GGGTGGATGAAAAGCACACAGAAGAACTTACGCTCGGCCGTTCCCACTCCAAATGATGACACTT  1450
Qy  1448  TGGATGAAGAGGAGTAGGGACTTGTGTCCATTCAAAACAATACCAACTGAGCATAATTCCAAGCCCTG  1507
Db  1451  TGGATGAAGGAGTAGGGACTTGTGTCCATTCAAAACAATACCAACTGAGCATAATTCCAAGCCCTG  1510
Qy  1508  CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAACTGAAAGCACAGCTGTCATTAGTA  1567
Db  1511  CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGAACTGAAAGCACAGCTGTCATTAGTA  1570
Qy  1568  ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGCATGGGGGT  1627
Db  1571  ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGCATGGGGGT  1630
Qy  1628  GGGGGTATGGGGAACAGTTGG  1647
Db  1631  GGGGGTATGGGGAACAGTTGG  1650
```

XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 240; 460pp; English.
XX
CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC2236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples given in comparison with
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

```
Qy   592 ------------------------------------------------------------GGGCTTCGTCGTGGCAC 607
                                                                     ||||||||||||||||
Db   551 TAGCTCTGCTCTAAATTGGGCGCCAAAATGTCCCTTTAATTTTGAACGAGGGCTTCGTCGTGGGAC 610

Qy   608 ATATGCAACAATGCCTATGATTTTACAAGCCTGATATGCTATTCTGAATATCCTATAGTAG 667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 ATATGCAACAATGCCTATGATTTTACAAGCCTGATATGCTATTCTGAATATCCTATAGTAG 670

Qy   668 ATGGGAAAAACTCTCCAATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 ATGGGAAAAACTCTCCAATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 730

Qy   728 GCAAAAAAGATCCATGCCCAGTGGCAGAAAAGAAGGGAAAATGATAAAGATTTTTACCTTGAATG 787
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 GCAAAAAAGATCCATGCCCAGTGGCAGAAAAGAAGGGAAAATGATAAAGATTTTTACCTTGAATG 790

Qy   788 ATTTTGGCTTCATGATCTTTCACCACCATTATTGTAAAACTGGTTCAGAAATCTTCTAGCTC 847
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 ATTTTGGCTTCATGATCTTTCACCACCATTATTGTAAAACTGGTTCAGAAATCTTCTAGCTC 850

Qy   848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGTAAAAATAGTATTCTATAGTG 907
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGTAAAAATAGTATTCTATAGTG 910

Qy   908 GCCTGGAAGCCTTTGGGGATGTTAAAATTAGAAGAGACACCTACTTTGATAGAGATGTGGAGA 967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 GCCTGGAAGCCTTTGGGGATGTTAAAATTAGAAGAGACACCTACTTTGATAGAGATGTGGAGA 970

Qy   968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTACTTG 1027
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTACTTG 1030

Qy  1028 TATCAAAATCAAAAATGGAAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCATTCGTTC 1087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1031 TATCAAAATCAAAAATGGAAAATATGTACACATCTTCAGTTATATGGTTCCCTTGCATTCGTTC 1090

Qy  1088 TAGCACAGTACTCACCTCACCTCACGCAAATTAGCAGGGAAGAGGAATTGGAGTGTTTTCTTAGGGTT 1147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1091 TAGCACAGTACTCACCTCACCTCACGCAAATTAGCAGGGAAGAGGAATTGGAGTGTTTTCTTAGGGTT 1150

Qy  1148 CTGGTTTGGCGTGCCACTCTGTACTCTCTTAAAGTCACACAGGAGAATGCTACACGGGGTCTG 1207
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1151 CTGGTTTGGCGTGCCACTCTGTACTCTCTTAAAGTCACACAGGAGAATGCTACACGGGGTCTG 1210

Qy  1208 CTCTTGATAAAATAACAGACAAGTTTATGTGATCTTAAAATCAAAGGCTTGATTCAAGAACTG 1267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211 CTCTTGATAAAATAACAGACAAGTTTATGTGATCTTAAAATCAAAGGCTTGATTCAAGAACTG 1270

Qy  1268 GTGTGGCACAGATGTCTTCGCTGAAAAACATGAAGTCAGAGAAGGACACCCATCATTTGG 1327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1271 GTGTGGCACAGATGTCTTCGCTGAAAAACATGAAGTCAGAGAAGGACACCCATCATTTGG 1330

Qy  1328 TCAACTATATTCCCCAGGGGTTCAAATAGAATTCACTCTTCTTTGAAGGAACAGTGGTTACTTAGTTA 1387
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1331 TCAACTATATTCCCCAGGGGTTCAAATAGAATTCACTCTTCTTTGAAGGAACAGTGGTTACTTAGTTA 1390

Qy  1388 GGGTGGATGGAGTAGGACTTGTGCATTCAAACATAGCCAACTGAGCATAATTCCAAGCCCTG 1447
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1391 GGGTGGATGGAGTAGGACTTGTGCATTCAAACATAGCCAACTGAGCATAATTCCAAGCCCTG 1450

Qy  1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATAATTCCAAGCCCTG 1507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1451 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATAATTCCAAGCCCTG 1510

Qy  1508 CCAAGAAAAGTACCAAGACTCCCTGCCACAGCGAGCAGAAACCTGAAGCAGCAGCTTCATTAGTTA 1567
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1511 CCAAGAAAAGTACCAAGACTCCCTGCCACAGCGAGCAGAAACCTGAAGCAGCAGCTTCATTAGTTA 1570

Qy  1568 ATGGGGAACATTAAGATACTCTGTGAGGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT 1627
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1571 ATGGGGAACATTAAGATACTCTGTGAGGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGT 1630

Qy   592 CAGGAGAGATATTGCTGTATATGCCACAGGAAAATGCTAGAGACCTACAGGTGGAGTTGGAGCAG 550
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   491 CAGGAGAGATATTGCTGTATATGCCACAGGAAAATGCTAGAGACCTACAGGTGGAGTTGGAGCAG 550
```

(additional alignment rows, coordinates:)
Qy 554 … 591 / Db 431 …
Qy 371 … 430
Qy 311 … 370
Qy 251 … 310
Qy 191 … 250
Qy 131 … 190
Qy 71 … 130
Qy 11 … 70

```
Qy   1628 GGGGGTATGGGAACAGTTGG 1647
          |||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650
```

RESULT 8

ACC62330

ID   ACC62330 standard; cDNA; 1650 BP.
XX
AC   ACC62330;
XX
DT   23-JUN-2003   (first entry)
XX
DE   Human NOV41e encoding cDNA SEQ ID NO:189.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0326236P.
PR   17-SEP-2001; 2001US-0326366P.
PR   17-SEP-2001; 2001US-0327781P.
PR   17-SEP-2001; 2001US-0328816P.
PR   17-SEP-2001; 2001US-0328817P.
PR   19-SEP-2001; 2001US-0323199P.
PR   17-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0249090P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-0236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML,   Alsobrook JP,   Anderson DW,   Berghs C,   Boldog FL;
PI   Burgess CE,   Casman SJ,   Catterton E,   Chant JS,   Chaudhuri A;
PI   Crabtree J,   Dipippo VA,   Edinger SR,   Eisen AJ,   Ellerman K;
PI   Gangolli EA,   Gerlach VL,   Giot L,   Gorman L,   Guo X,   Gusev VV,   Ji W;
PI   Kekuda R,   Khramtsov NV,   Lepley DM,   Li L,   Liu X;
PI   Malyankar UM,   Miller CM,   Ooi CE,   Ort T,   Padigaru M,   Patturajan M;
PI   Pena CEA,   Rieger DK,   Rothenberg ME,   Shenoy SG,   Shimkets RA;
PI   Spaderna SK,   Spytek KA,   Taupier RJ,   Twomlow N,   Vernet CAM,   Voss EZ;
PI   Zerhusen BD,   Zhong M;
```

XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54261.
XX
XX   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 242; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

```
Query Match             68.4%;  Score 1370;  DB 7;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 4,1e-259;
Matches 1509;   Conservative   0;   Mismatches   5;   Indels 126;   Gaps   1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAAATGCAGAAGCTTGCTGGCCAAAAGATG 193
          || ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAAATGCAGAAGCTTGCTGGCCAAAAGATG  70

Qy   194 TGGGAATTGTTGCCCTTGAGATCTTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71 TTGGGAATTGTTGCCCTTGAGATCTTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAAATATGATGGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAAATATGAAGGGTCTTAGATGCTGGGAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
          |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 250

Qy   374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGGAAGTTGGAACAGAGACA 433
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGGAAGTTGGAACAGAGACA 310

Qy   434 TCATCGACAAATCAAAGTCTTGTGAAGACTAAATTTGATGCAGCTGTTTTGAAGAGTCTGGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTTGTGAAGACTAAATTTGATGCAGCTGTTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAATCGACACACAACTAGTATGGAGGGCACAGCGTGCCTGTGTCT 553
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   371 ATACAGATATAGAAGGAATCGACACACAACTAGTATGGAGGGCACAGCTGTGCCTGTGTCT 430

Qy   554 TCAATGCTGTTAACTGGAGTTGAGTCCAGCTCTTGGGAT---------------------- 591
          |||||||||||||||||||||||||||||||||||||||
Db   431 TCAATGCTGTTAACTGGAGTTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTAGTTG 490
```

```
Qy    592 ------------------------------------------------------------ 591

Qy    592 ------------------------------------------------------------ 591
Db    551 TAGCTCTCTGCTATTGGGCCAAAATGCTTCTCTTTAATTTTTGACAGGAAATGCTAGACAG 610
                                                     --GGGCTTCGTGGGGACAC
Db    491 CAGGAGATATTGCTGTATATGCCACAGGAAATGCTAGAGCCTACAGGTGGGAGTTGGAGCAG 550

Qy    592 ------------------------------------------------------------ 591
Db    551 TAGCTCTCTGCTATTGGGCCAAAATGCTTCTCTTTAATTTTTGACAGGAAATGCTAGACAG 610
                                                     --GGGCTTCGTGGGGACAC 607

Qy    608 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATCTCTATAGTAG 667
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATCTCTATAGTAG 670

Qy    668 ATGGGAAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCGCTGCTATTCTGTCTACT 727
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGGAAAACTCTCCATACAGTGCTACCCTCAGTGCATTAGACCGCTGCTATTCTGTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCCAGAGAAAGAGGGAAAATGATAAAGATTTTTACCTTGAATG 787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCATGCCCAGTGGCCAGAGAAAGAGGGAAAATGATAAAGATTTTTACCTTGAATG 790

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATTAGTATAAACTGGTTCAGAAAATCTCTAGCTC 847
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTGGCTTCATGATCTTTCACTCACCATTAGTATAAACTGGTTCAGAAAATCTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATACTATCTCTATAGTG 907
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATACTATCTCTATAGTG 910

Qy    908 GCCTGGGAAGCCTTTGGGGATGTTAAATTAGAAGACCTACTTTGATAGAGGATGTGGAGA 967
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGGAAGCCTTTGGGGATGTTAAATTAGAAGACCTACTTTGATAGAGGATGTGGAGA 970

Qy    968 AGGCATTTATGAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGCCATCTTTTACTTG 1027
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGCCATCTTTTACTTG 1030

Qy   1028 TATCAAAATCAAAAATGGAAATATGTACACATCTTCAGTATATAGTTGGTTCCCTTGCATCTGTTC 1087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAAATCAAAAATGGAAATATGTACACATCTTCAGTATATAGTTGGTTCCCTTGCATCTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTT 1147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGCCTCCACTCGTACTCTTAAAGTCACAAGAGATGCTACACCGGGGTCTG 1207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGCCTCCACTCGTACTCTTAAAGTCACAAGAGATGCTACACCGGGGTCTG 1210

Qy   1208 CTCTTGATAAAAATAAACACACAAGAGTTTATGTGATCTTTGATTCAAGGCTTGATCAGAACTG 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAAATAAACACACAAGAGTTTATGTGATCTTTGATTCAAGGCTTGATCAGAACTG 1270

Qy   1268 GTGTGGCACCGATGTCTTCGCGTCTGAAAAACATGAAGCTCAGAGGACACCCATCATTGG 1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCGATGTCTTCGCGTCTGAAAAACATGAAGCTCAGAGGACACCCATCATTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGGATGAAAAAGCAGAAAGAACTTACGGCTCGCGCTCCCACTCCCAAAATGATGACACTT 1447
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGGATGAAAAAGCAGAAAGAACTTACGGCTCGCGCTCCCACTCCCAAAATGATGACACTT 1450

Qy   1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1510

Qy   1508 CCAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCTGAGCAGCTGTCATTAGTTA 1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCTGAGCAGCTGTCATTAGTTA 1570

Qy   1568 ATGGGGACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGTGGGCCATGGGGGT 1627
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGTGGGCCATGGGGGT 1630

Qy   1628 GGGGGTATGGGAACAGTTGG 1647
          ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650
```

```
PI    Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA,
PI    Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI    Zerhusen BD, Zhong M;
XX
DR    WPI; 2003-313241/30.
DR    P-PSDB; ABR54269.
XX
PT    Novel human proteins and nucleic acid encoding the proteins, useful for
PT    diagnosis, treatment and prevention of disorders involving the human
PT    protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS    Claim 20; Page 246; 460pp; English.
XX
CC    The present invention describes isolated human NOVX proteins, where X is
CC    1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC    ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC    hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC    antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC    antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC    antiparkinsonian and antilipaemic activities, and can be used in gene
CC    therapy. NOVX proteins are useful for treating or preventing a pathology
CC    associated with a NOVX protein in humans and for treating a syndrome
CC    associated with the human disease. NOVX nucleic acids, proteins and
CC    antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC    atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC    valve disease, tuberous sclerosis, scleroderma, diabetes, metabolic
CC    disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC    congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC    hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC    disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC    infectious disease, anorexia, cancer-associated cachexia, cancer,
CC    Alzheimer's disease, Parkinson's disease, immune disorders,
CC    haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC    ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC    sequences, which are used in examples from the present invention.
CC    ABR54277 represents a human trypsinogen protein given in comparison with
CC    the human NOV35b protein in the exemplification of the present invention
XX
SQ    Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match          68.4%;  Score 1370;  DB 7;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 4.1e-259;
Matches 1509;  Conservative   0;  Mismatches   5;  Indels 126;  Gaps   1;

Qy      134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
            ||||||||||||| ||||||| |||||||||||||||||||||||||||||||||||||||
Db       11 GCTCTTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy      194 TGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATAGTTTGATCAAGCAGAGTTGG 253
            | ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db       71 TTGGGAATTGTTGCCCTTGAGATCTATTTTCTTCTCAATATAGTTTGATCAAGCAGAGTTGG 130

Qy      254 AAAAATATGATGGGTGTAGAGTCTTGGGAAAGTATACACATTGGCTTGGGCTCCAAGAGG 313
            |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db      131 AAAAATATGATGGGTGTAGAGTCTTGGGAAGTATACACATTGGCTTGGGCTCCAAGATGG 190

Qy      314 GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGCAATGGTTGGTTCAGAATCTTA 373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      191 GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGCAATGGTTGGTTCAGAATCTTA 250

Qy      374 TGGAGAGAAATAAACTTTCCTATCATTGGGCGGCGGCTGGGAGTTGGAACAGAGACAA 433
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      251 TGGAGAGAAATAAACTTTCCTATCATTGGGCGGCGGCTGGGAGTTGGAACAGAGACAA 310

Qy      434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 370

Qy      494 ATACAGATATAGAAGGAAGAATCGACACAACTAATCATGCTATGGAGGCACAGCTTGCTGTCT 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      371 ATACAGATATAGAAGGAAGAATCGACACAACTAATCATGCTATGGAGGCACAGCTTGCTGTCT 430

Qy      554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------------ 591
            ||||||||||||||||||||||||||||||||||||||
Db      431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490

Db      490 ----------------------------------------------------------- 591
            
Qy      592 CAGGAGGATATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGGAGTGGAGCAG 551
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      491 CAGGAGGATATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGGAGTGGAGCAG 550

Qy      608 ATATGCAACAATGCCTATGATTTTTACAAGGCCTGATAATGCCTATCCTATATCCTATAGTAG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 610

Qy      608 ATATGCAACAATGCCTATGATTTTTACAAGGCCTGATAATGCCTATCCTATATCCTATAGTAG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 610

Db      607 -----------GGGCTTCGTGGGACAC 607

Db      551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC 610

Qy      608 ATATGCAACAATGCCTATGATTTTTACAAGGCCTGATAATGCCTATCCTATATCCTATAGTAG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      611 ATATGCAACAATGCCTATGATTTTTACAAGGCCTGATAATGCCTATCCTATATCCTATAGTAG 670

Qy      668 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACT 727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      671 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACT 730

Qy      728 GCAAAAAGATCCATGCCCAGTGCGCAGAAAGAGGGAAAATGATAAAAGATTTTACTTGAATG 787
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      731 GCAAAAAGATCCATGCCCAGTGCGCAGAAAGAGGGAAAATGATAAAAGATTTTACCTTGAATG 790

Qy      788 ATTTTCGGCTTCATGATCTTTCACTCACCATATTGTAAACGGGTTCAGAAATTCTCTAGCTC 847
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      791 ATTTTCGGCTTCATGATCTTTCACTCACCATATTGTAAACGGGTTCAGAAATTCTCTAGCTC 850

Qy      848 GGATGTTGCTGAATGACACTTCCTTAATGACCAGAATAGAGATRAAAATAGTAGTCTATAGTG 907
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      851 GGATGTTGCTGAATGACACTTCCTTAATGACCAGAATAGAGATAAAAATAGTAGTCTATAGTG 910

Qy      908 GCCTCGGAAGGCCTTTGGGGATGTTAAATTAGGAGAGACACCTACTTTGATAGGATGTGGAGA 967
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      911 GCCTCGGAAGGCCTTTGGGGATGTTAAATTAGGAGAGACACCTACTTTGATAGGATGTGGAGA 970

Qy      968 AGGCATTTATGAAGGCTAGCTTCAGTGCAAAACAAAGGCATCTTTACTTG 1027
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      971 AGGCATTTATGAAGGCTAGCTTCAGTGCAAAACAAAGGCATCTTTACTTG 1030

Qy     1028 TATCAAAATCAAAAATGGAAATATGTACACATCTTCAGTATGATATGGTTCCCTGCATCGTGTC 1087
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1031 TATCAAAATCAAAAATGGAAATATGTACACATCTTCAGTATGATATGGTTCCCTGCATCGTGTC 1090

Qy     1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAGAATTGGGAGTGTTTTCTTATGGTT 1147
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAAGAATTGGGAGTGTTTTCTTATGGTT 1150

Qy     1148 CTGGTTTGGGTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGGGGAGGAATGCTACACGGG 1207
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1151 CTGGTTTGGGTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGGGGAGGAATGCTACACGGG 1210

Qy     1208 CTCTTGATAAAAATAACAGCAAAGTTTATGTGATCTTAAATCAAGGCTTGATTGATTCAAGAACTG 1267
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1211 CTCTTGATAAAAATAACAGCAAAGTTTATGTGATCTTAAATCAAGGCTTGATTGATTCAAGAACTG 1270

Qy     1268 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTTCAGAGGAACACCCCATCATTGG 1327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1271 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTTCAGAGGAACACCCCATCATTGG 1330

Qy     1328 TCAACTATATTCCCAGGGTTCAATAGATTCACTCTTTGAAGGAGGGACGTGGTACTTAGTTA 1387
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1331 TCAACTATATTCCCAGGGTTCAATAGATTCACTCTTTGAAGGAGGGACGTGGTACTTAGTTA 1390

Qy     1388 GGGTGGCATGAGAAAAGCACAGAAGAGAACTTACGTCGGTCCCAACTCCAAATGATGACACTT 1447
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1391 GGGTGGCATGAGAAAAGCACAGAAGAGAACTTACGTCGGTCCCAACTCCAAATGATGACACTT 1450

Qy     1448 TGGATGAAGGAGTAGGACTTGTCGCATTGCAATTCAAAACATTTCCAAGCCCTG 1507
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1451 TGGATGAAGGAGTAGGACTTGTCGCATTGCAATTCAAAACATTAGCACGCCCTG 1507

Qy     1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCACAGAAACTTGAACGCAGCTGTCATTAGTA 1567
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1451 TGGATGAAGGAGTAGGACTTGTCGCATTGCAATTCAAAACATTAGCACGCCCTG 1510
```

```
Db   1511 CCAGAAAGTACCAAGACTCCCTGCCAGCAGCAGAGAACCTGAAGCAGCAGCTGTCATTAGTA 1570
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1568 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGTGTGGGTGGGGTCATGGGGT 1627

Db   1571 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGTGTGGGTGGGGTCATGGGGT 1630
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1628 GGGGGTATGGGAACAGTTGG 1647
              ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650

RESULT 10
ACC62334
ID   ACC62334 standard; cDNA; 1650 BP.
XX
AC   ACC62334;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human NOV41l encoding cDNA SEQ ID NO:197.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflamatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; Alzheimer's disease; Parkinson's disease;
KW   cancer-associated cachexia; haematopoietic disorder; dyslipidaemia;
KW   immune disorder; metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002;  2002WO-US028538.
XX
PI   07-SEP-2001; 2001US-0318120P.
PI   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   17-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0316663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393333P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
```

```
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ott T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zernusen BD, Zhong M;
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 244; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, anorexia, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match             68.4%;  Score 1370; DB 7;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 4.1e-259;
Matches 1509; Conservative  0; Mismatches  5; Indels 126; Gaps  1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
             ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db    11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy   194 TGGGAAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
             ||||||||||||||||||||||||| |||||||| ||||||| |||||||||||||||||
Db    71 TTGGGATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAATATGATGGTGTTAGATGCTGGAAAGTAATATACCATTGGCTTGGGCCAGGCCAAGATGG 313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAATATGATGGTGTTAGATGCTGGAAAGTATTACCATTGGCTTGGGCCAGGCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAGATATTAACTCTCTTCTTTGCATGACTGGTTGCATGAGAATCTTA 373
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAGATATTAACTCTCTTCTTTGCATGACTGGTTGCATGAGAATCTTA 250

Qy   374 TGGGAGGAAAATAACCTTTCCTTCTTATGTATTGCATTGGCCGGCGTTGGGACAGTTGGAACACAA 433
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGGAGGAAAATAACCTTTCCTTCTTATGTATTGCATTGGGCGGCGTTGGGAAGACAGAGACAA 310

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTGAATGCAGCTCGTTTGAAGAGTCTGGGA 493
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTGAATGCAGCTCGTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAAATCGACACAACTAATGCATGCTAATGGAGGCACACAGCTGTCT 553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   494 ...
```

```
Db   371  ATACAGGATATAGAGAGGAAATCGACACAACTAATGCATGCTATGGAGGAGGCACAGCTGCTGTCT                            430
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   554  TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----------------------------                            591

Qy   431  TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT                                                         490
          |||||||||||||||||||||||||||||||||||||
Db        ----------------------------------------                                                      591

Qy   592  CAGGAGGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGTTGGAGCAG                                 550
                                                  ---------------GGGCTTGCTGGGGACAC                    607

Qy   592                                          ---------------GGGCTTGCTGGGGACAC                    607

Db   551  TAGCTCTTGCTAATTGGGGCCAAAATGCTCTCTTTAATTTTTTGAACGAGAGGGCTTCGTGTGGGACAC                          610

Qy   608  ATATGCAACATGCCCAGTGCATGAATTTTACAAGCCTGATATGCTATCTGAATAATCCTATAGTAG                             667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611  ATATGCAACATGCCCAGTTAATGAATTTTTACAAGCCCTGATATGCTAACTGAATAATCCTATAGTAG                           670

Db   668  ATGGAAAAACTCTCCAATACAGTGCTACCTCAGTGCATTAGACCCGCTGCTACTTCGTGTCTACT                              727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   671  ATGGGAAAACTCTCCAATACAGTGCTACCTCAGTGCATTAGACCCGCTGCTACTTCGTGTCTACT                              730

Qy   728  GCAAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATCGATAAAGATTTTACCTTGAAATG                                787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731  GCAAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAAATG                                 790

Db   788  ATTTTGGCTTCAATGAATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC                                 847
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   791  ATTTTGGCTTCAATGAATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC                                 850

Qy   848  GGATGTTGCTGAATGAACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTG                                  907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851  GGATGTTGCTGAATGAACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTG                                  910

Db   908  GCCTGGAAGCCCTTGGGGGATGTTAAATTAGAGAAGACACCTACTTTGATAGAGAATGTGGAGA                               967
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   911  GCCTGGAAGCCCTTGGGGGATGTTAAATTAGAGAAGACACCTACTTTGATAGAGAATGTGGAGA                               970

Qy   968  AGGCATTTATGAGAGGCTAGCTCTCGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG                                 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971  AGGCATTTATGAGAGGCTAGCTCTCGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG                                 1030

Db   1028 TATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC                                   1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1031 TATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC                                   1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAGGAGAATTGGAGTGTTTTCTTATGGTT                                   1147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAGGAGAATTGGAGTGTTTTCTTATGGTT                                   1150

Db   1148 CTGGTTTGGCTGCCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG                                  1207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1151 CTGGTTTGGCTGCCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG                                  1210

Qy   1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAAACTG                                  1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAAACTG                                  1270

Db   1268 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCATCATTGG                                    1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1271 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCATCATTGG                                    1330

Qy   1328 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAGCGTGGTACTTAGTTA                                  1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAGCGTGGTACTTAGTTA                                  1390

Db   1388 GGGTGGGATGAAAAGCACAGAAGAAGAACTTAGCGTCGGGCCTCCCACTCCAAATGATGACACTT                              1447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1391 GGGTGGGATGAAAAGCACAGAAGAAGAACTTAGCGTCGGGCCTCCCACTCCAAATGATGACACTT                              1450

Qy   1448 TGGATGAAGGAGGTAGGAGACTTGTGTGCATTCAAACATAGCAACTGAGCATTAATTCCAAGCCCTG                            1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGGTAGGAGACTTGTGTGCATTCAAACATAGCAACTGAGCATTAATTCCAAGCCCTG                            1510
```

Qy   1508 CCAAGAAAGTACCAAGACTCCCTCCCAGCAGCCAGAACCCTGAAGCAGCTGTCTCATTAGTA   1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCTCCCAGCAGCCAGAACCCTGAAGCAGCTGTCATTAGTA   1570

Qy   1568 ATGGGGAACATTAAGATACTTCTTGTCAAGGTGCAAGAACTTCAGGGTGGTGGGGTGGGGCATGGGT   1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTTCTTGTGAAGGTGCAAGAACTTCAGGGTGGTGGGGTGGGGCATGGGT   1630

Qy   1628 GGGGGTATGGGAACAGTTGG   1647
          |||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG   1650

RESULT 11
ACC62327
ID   ACC62327 standard; cDNA; 1650 BP.
XX
AC   ACC62327;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV41b encoding cDNA SEQ ID NO:183.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0328216P.
PR   17-SEP-2001; 2001US-0328217P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0349696P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0249990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   17-MAY-2002; 2002US-0377908P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX

PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SU, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman VJ, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Liu L, Ji W;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M,
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54258.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 240-241; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62245 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antiischaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match   68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity   92.0%;
Matches 1509;   Conservative   0; Pred. No. 4.1e-259;
                                     Mismatches 5;   Indels 126; Gaps 1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAGATG 193
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAGATG 70

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 TTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAATATGATGGTAGATGCTTGGAAAGTATACCATTGGCTTGGGCTCCAAGGCCAAGATGG 313
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAATATGATGGTAGATGCTTGGAAAGTATACCATTGGCTTGGGCTCCAAGGCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGCATGATGACGTGGTTCAGATCTTA 373
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 250

Qy   374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGCGCTGGAAGTTGGAACAGAGACAA 433
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGTTCAATAGATTCAGTACAGAGACA 310

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCCAGCTGTTTGAAGAGTCTGGGA 493
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCCAGCTGTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGAGATATAGAAGGAAATCGACACACAACTAAATGCATTGCTATGGAGGCACAGCTGCTGTCT 553
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 ATACAGAGATATAGAAGGAAATCGACACACAACTAAATGCATTGCTATGGAGGCACAGCTGCTGTCT 430

Qy   554 TCAATGCTCGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------ 591
               ||||||||||||||||||||||||||||||||||||||||
Db   431 TCAATGCTCGTTAACTGGATTGAGTCCAGCTCTTGGGGATTGGACACGGTGTATGCCCTGGGTAGTTG 490

Qy   592 --- 591
Db   491 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTACACAGGTGGAGTGGAGCAG 550

Qy   592 ----------------------GGGCTTCGGTGGGAGCAC 607
                                      |||||||||||||||
Db   551 TAGCTCGTGCTTAATTGGGGCCAAAATGCTTCCTTTAATTTTTGAACGAGGGGCTTCGGTGTGGGACAC 610

Qy   608 ATATGCAACATGCCTATGATTTTTTACAAGCCTGATATGCTATTCTGAATATATCCTATAGTAG 667
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 ATATGCAACATGCCTATGATTTTTTACAAGCCTGATATGCTATTCTGAATATATCCTATAGTAG 670

Qy   668 ATGGAAAAACTCTCATACAGTGCTACCTCAGTGCATTAGAACGCTGCTACTTCTGTCTACT 727
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 ATGGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAACGGCGTCGCTATTCTGTCTACT 730

Qy   728 GCAAAAAGATCCATCCCACTGGCAGAAGAGGGGGAAAATGATAAGCAATTTACCTTGAATG 787
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 GCAAAAAGATCCATCCCACTGGCCAGAAGAGAAAGTGATAAGAGATTTTACCTTGAATG 790

Qy   788 ATTTTGGGCTTCATGATCTTTCACTCACCCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 847
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 ATTTTGGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 850

Qy   848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATTAGAGATAAAAATAGTAGTTCTTATG 907
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAATAGTAGTTCTTATG 910

Qy   908 GCCTGGAAGCCTTGGGGATGTTAAAATTAGAAGACCACCTACTTTGATAGAGATGTGGAGA 967
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 GCCTGGAAGCCTTGGGGATGTTAAAATTAGAAGACCACCTACTTTGATAGAGATGTGGAGA 970

Qy   968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGGCATCTTTACTTG 1027
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGGCATCTTTACTTG 1030

Qy   1028 TATCAAAATCAAAAATGGAAATATGTACACATTCTTCAGTGTATATGGTTCCCTTGCATCGTGTTC 1087
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAAATCAAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCGTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCCAATTAGCAGGGAAGGAGAAGAATTGGAGTGTTTTTCTTATGGTT 1147
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCCAATTAGCAGGGAAGGAGAAGAATTGGAGTGTTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGCTGCGAAGCTACTCTGCTTCAAGTCACACAAGATGCTACACGGGGGTCTG 1207
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGGCTGCCAACTCTGTACTCTGCTTAAAGTCACACAAGATGCTACACGGGGGTCTG 1210

Qy   1208 CTCTTGATAAAATAACACAGAGTATATTGAATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATTGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGGCACCAGATGTCTTCGGCTGAAAACATGAGAGCTCAGAGAGAGCACCCATCATTGG 1327
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGGCACCAGATGTCTTCGGCTGAAAACATGAGAGCTCAGAGAGAGCACCCATCATTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGGTTCAATAGATTCACTCTTTGAAGGAGACGTGGTACTTAGTTA 1387
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAGACGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTT 1447
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTT 1450

Qy 1448 TGGATGAAGAGGAGTAGGACTTGTGCATTCAAACATAGCATATTCCAAGCCCTG 1507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1451 TGGATGAAGAGGAGTAGGACTTGTGCATTCAAACATACCAACTGAGCATATTCCAAGCCCTG 1510

Qy 1508 CCAAGAAAGTACCAAGACTTCCCTGCCACAGCAGACTGAGCATATTCCAAGCCCTG 1567
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1511 CCAAGAAAGTACCAAGACTCCCTGCCACACGACCAGAACCTGAACGCAGCTGTCATTAGTA 1570

Qy 1568 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGGTGGGGGGT 1627
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1571 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGT 1630

Qy 1628 GGGGGGTTATGGGAACAGTTGG 1647
         ||||||||||||||||||||||
Db 1631 GGGGGTTATGGGAACAGTTGG 1650

RESULT 12
ACC62332
ID ACC62332 standard; cDNA; 1650 BP.
XX
AC ACC62332;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV41g encoding cDNA SEQ ID NO:193.
XX
KW Human; NOVX; antiatherosclerotic; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200302301-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.
XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0322656P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0328816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323636P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0341144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0388663P.
PR 02-JUL-2002; 2002US-0393332P.
PR 17-JUL-2002; 2002US-0396412P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-0025417.
XX
XX
PA (CURA-) CURAGEN CORP.
XX
PA Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ort T, Padigaru M, Patturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX
DR WPI; 2003-313241/30.
DR P-PSDB; ABR54263.
XX
PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS Claim 20; Page 243; 460pp; English.
XX
CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOV protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX
SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match 68.4%; Score 1370; DB 7; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

Qy 134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTTGGCCAAAAGATG 193
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAAGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy 194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 253
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 71 TTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 130

Qy 254 AAAAAATATGATGGGTGTAGATGCTTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 131 AAAAAATATGATGGGTGTAGATGCTTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 190

Qy 314 GCTTCTCTGCACAGATAGAGAAAGATATTAAACTCTCTTTTGCAGTGACTGTGGTTCAGATCTTA 373
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191 GCTTCTCTGCACAGATAGAGAAGATATTAAACTCTCTTTTGCAGTGACTGTGGTTCAGATCTTA 250

Qy 374 TGGAGAGAAAATAAACCTTTCCTATGATTTGCATTGGGGCGGCGGCTGGAACAGAGACA 433
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 251 TGGAGAGAAAATAAACCTTTCCTATGATTTGCATTGGGGCGGCGGCTGGAACAGTTGGAACAGAGACA 310

```
Qy    434 TCATCGACAAATCAAAGTCTGTGAAGAGTAAATTGATCGACGTGTTGTCAGCTCTGGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCATCGACAAATCAAAGTCTGTGAAGAGTAAATTGATCGACGTGTTTGAAGAGTCTGGGA 370

Qy    494 ATACAGATAATAGAAGGAATCGACAACAACTAATCGCATCGCTATGCAGCTGCTGCTCT 553
          ||||||||||||||||||||||||||||||||||||||||||||
Db    371 ATACAGATTATAGAAGGAATCGACAACAACTAATCGCATGCTATGGACAGCTGCTGCTGCT 430

Qy    554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------ 591
          ||||||||||||||||||||||||||||||
Db    431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAGGACCGGTATGCCCTGGTAGTTG 490

Qy    592 ------------------------------------------ 591
Db    491 CAGGAGATATTGCTGTATATGCCACACGGAAATGCTAGACCTACACGGTGGAGTGGAGCAG 550

Qy    592 ------------------------- 591
Db    551 TAGCTCTGCTAATTGGGGCCAAAATGCTCCTTTAATTTTTGAACCAGAGGCTTCGTGGGACAC 610

Qy    608 ATATGCAACACATGCCTATGATTTTTACAAGCCTGATAGCTATCGTCGTTATGGAGTAG 667
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACACATGCCTATGATTTTTACAAGCCTGATATGCTATCGTCCTATAGGTAG 670

Qy    668 ATGGAAAAACTCTCCATACAGTGCTACCCTCAGTGCCTGCATTAGACCCGCTGCTATTCTGTCTACT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGGAAAACTCTCCATACAGTGCTACCCTCAGTGCCTGCATTAGACCCGCTGTATTCTGTTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGAAGAATGATAAAGGATTTTACCTTGAATG 787
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGAAGAATGATAAAGGATTTTACCTTGAATG 790

Qy    788 ATTTGGCTTCATGATCTTTCACTCACCATATGTGTAAACTGGTTCAGAAAATCTTCTAGCTC 847
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTGGCTTCATGATCTTTCACTCACCATATGTGTAAACTGGTTCAGAAAATCTTCAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATTAGAGATAAAAAATAGTATCTATAGTG 907
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATTAGAGATAAAAAATAGTATCTATAGTG 910

Qy    908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAGCACCTACTTCTTTGATAGGAGATGTGGAGA 967
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAGCACCTACTTCTTTGATAGAGATGTGGAGA 970

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCCTTAAATGACCAAACAAAGGCATCTTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGTCAGAAAACAAAGGCATCTTTACTTG 1030

Qy   1028 TATCAAATCAAAATGGAAATATATCTACACATCTCAGTATATGTGTTCCCTTGCATCTGTTC 1087
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAATCAAAATGGAAATATATCTACACATCTCAGTATATGTGTTCCCTTGCATCTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAAGAGAATTGGAGTGGTTTTCTTATGGTT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAAGAGAATTGGAGTGGTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGCTGCCACTCTGTACTCTGTTAAAGTCACAACGCCCGGGGTCTG 1207
          ||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGCTGCCACTCTGTACTCTGTTAAAGTCACAACGATGCTACACCGGGGGTCTG 1210

Qy   1208 CTCTTGATAAAATAACACACAGTTTATGTGATCTTGATTCAATTCAAGAGACTG 1267
          ||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACACACAGTTTATGTGATCTTAAATCAAGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGCACCGGATGTCTTCGCTGGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCGGATGTCTTCGCTGGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGGTTCAATAGATTCACTCTTTGAAGGAAACGTGGTACTTAGTTA 1387
          ||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAACGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGTCCACTCCAAAATGATGACACTT 1447
          ||||||||||||||||||||||||||||||||||||||||||
Db   1388 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGTCCACTCCAAAATGATGACACTT 1447
```

```
Db   1391 GGGTGGATGAAGAGAGTAGGAGCACAGAAGAAGAACTTACGCTCGGCGTCCCACTCCAAAATGATGACACTT 1450

Qy   1448 TGGATGAAGGAGAGTAGGGACTTGTTGCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
          ||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGAGTAGGGACTTGTTGCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTG 1510

Qy   1508 CCAAGAAAGTACCAAGACTCCCCTGCCACAGCAGCAGGAACCTGAAGCAGCTTGTCATTAGTA 1567
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCCTGCCACAGCAGCAGGAACCTGAAGCAGCTTGTCATTAGTA 1570

Qy   1568 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGGTGGGGTGGGGCATGGGGT 1627
          ||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGGTGGGGTGGGGCATGGGGT 1630

Qy   1628 GGGGGTATGGGAACAGTTGG 1647
          ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650
```

RESULT 13

```
ADE38394

ID    ADE38394 standard; DNA; 1650 BP.
XX
AC    ADE38394;
XX
DT    29-JAN-2004  (first entry)
XX
DE    Human protein 9389 gene sequence.
XX
KW    tumourigenic disorder; angiogenic disorder; aberrant gene expression;
KW    aberrant protein activity; cytostatic; antithyroid; antidiabetic;
KW    ophthalmological; cancer; breast cancer; colon cancer; lung cancer;
KW    prostatic cancer; Grave's disease; diabetic retinopathy; gene; ds;
KW    protein 9389.
XX
OS    Homo sapiens.
XX
FH
FT    Key             Location/Qualifiers
FT    CDS             22..1584
FT                    /tag= a
FT                    /product= "Human protein 9389"
XX
PN    WO2003065006-A2.
XX
PD    07-AUG-2003.
XX
PF    30-JAN-2003;  2003WO-US002588.
XX
PR    31-JAN-2002;  2002US-0353600P.
PR    15-MAR-2002;  2002US-0364517P.
PR    09-APR-2002;  2002US-0371075P.
PR    10-APR-2002;  2002US-0371507P.
PR    16-APR-2002;  2002US-0372984P.
PR    19-APR-2002;  2002US-0374194P.
PR    24-MAY-2002;  2002US-0382995P.
PR    31-MAY-2002;  2002US-0385023P.
PR    14-JUN-2002;  2002US-0388853P.
PR    17-JUN-2002;  2002US-0389395P.
PR    25-JUN-2002;  2002US-0391324P.
PR    15-JUL-2002;  2002US-0395944P.
PR    22-JUL-2002;  2002US-0397726P.
PR    13-AUG-2002;  2002US-0403046P.
PR    22-AUG-2002;  2002US-0405155P.
PR    27-AUG-2002;  2002US-0406361P.
PR    25-OCT-2002;  2002US-0421195P.
PR    12-NOV-2002;  2002US-0425456P.
PR    19-NOV-2002;  2002US-0427626P.
PR    10-DEC-2002;  2002US-0431222P.
XX
PA    (MILL-) MILLENNIUM PHARM INC.
XX
PI    Hunter JJ, Macbeth KJ, Tsai F, Leseon A, Lightcap ES;
```

PI   Williamson MW, Rudolph-Owen LA;
XX
DR   WPI; 2003-646176/61.
DR   P-PSDB; ADE38395.
XX
PT   Treating subject having tumorigenic disorder or angiogenic disorder
PT   caused by aberrant polypeptide e.g., N-formylpeptide receptor or nucleic
PT   acid, by administering a modulator.
XX
PS   Disclosure; SEQ ID NO 55; 454pp; English.
XX
CC   This invention relates to a novel method of treating a human subject
CC   having a tumourigenic disorder or angiogenic disorder, caused by aberrant
CC   gene expression or activity of an isolated protein, by administering a
CC   modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC   or ophthalmological activity. The method is useful for treating a subject
CC   having a tumourigenic or angiogenic disorder, in particular for treating
CC   cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC   cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC   present sequence is a DNA sequence which encodes the novel isolated human
CC   protein 9389 of the invention.
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Query Match   68.4%; Score 1370; DB 9; Length 1650;
Best Local Similarity 92.0%; Pred. No. 4.1e-259;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

[Sequence alignment — Qy (query) vs Db (database); position numbers shown at left and right of each row]

```
Qy    134 GCTCTTTCACCATGCCTGGATCACTTCCTTTTGAATGCGAAGAGCTTGCTTGGGCAAAAGATG 193
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     11 GCTCTTTCACCATGCCTGGATCACTTCCTTTTGAATGCGAAGAGCTTGCTGGCCAAAAGATG  70

Qy    194 TGGGAATTGTTGCCCTTGAGATCTATTTTTCCTTCTCAAATATGTTGATCAGCAGAGTTGG 253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TTGGGAATTGTTGCCCTTGAGATCTATTTTTCCTTCTCAAATATGTTGATCAGCAGAGTTGG 130

Qy    254 AAAAAATATGATGCTGTGAGATGCTGGGAAAGTATACCATTGGCTTGGGACGTCCAAGATGG 313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 AAAAATATGATGGGTGTTAGATGCTGGGAAGTATACCATTGGCTTGGGACGTCCAAGATGG 190

Qy    314 GCTTCTCGCACAGATAGAGAAGAATATTAACTCTCTTTTGCATGACTGTGGTTCAGAATCTTA 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 GCTTCTGCACAGATAGAGAAGAATATTAACTCTCTTTTGCATGACTGTGGTTCAGAATCTTA 250

Qy    374 TGGAGAGAAAATAACCTTTCCTTCATTGGGGGGCGGCTGGGAGTTGGAACAGAGAACAA 433
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 TGGAGAGAAAATAACCTTTCCTATCATTGGGCGGCCTGGAAGTTGGAACAGAGAACAA 310

Qy    434 TCATCGACAAATCAAAAGTCTGTGAAGACTAAATTGATGCAGCTGTTTTGAAGAGTCTGGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCATCGACAAATCAAAAGTCTGTGAAGACTAAATTGATGCAGCTGTTTTGAAGAGTCTGGGA 370

Qy    494 ATACAGATATAGAAGGAAATCGACACACTAATGCAATGCATATGGGAGGGCACAGCTGCTGTCT 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 ATACAGATATAGAAGGAAATCGACACAACTAATCCTATGGGAGGGCACAGCTGCTGTCT 430

Qy    554 TCAATCGCTTGTTAACTGGGAATTGAGTCCAGCTCTTGGGAT------------------- 591
          |||||||||||||||||||||||||||||||||||||||||||
Db    431 TCAATCGCTTGTTAACTGGGAATTGAGTCCAGCTCTTGGGATGGGATGAGACGGTATGCCCTGGTAGTTG 490

Qy    592 -------------------------------------------------------------- 591
          
Db    491 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTACAGACCTACAGGTGGAGTTGGAGCAG 550

Qy    592 -------------------------------------------------------------- 591
          
Db    551 -----GGGCTTCGTGGGGACAC 607

Qy    608 ATATGCGAACATGCCTATAGATTTTTACAAGCCTGAATATCTTGAATATCCTTATTAGTAG 667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACATGCCTATGATTTTTACAAGCCTCGATATGCTATTCTGAATATCCCTATAGTAG 670

Qy    668 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCCATTAGACCGCTGCTACTCTGTCTACT 727
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGGAAAACTCTCCATACAGTGCTACCTCAGTGCCATTAGACCGCTGCTACTCTGTCTACT 730
```

```
Db    671 ATGGGAAAACTCTCCATACAGTGCCAGAAAGAAGAGGGAAAATGATTAGAGATTTTTACCTTGAATG 730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAAGAGGGAAAATGATTAGAGATTTTTACCTTGAATG 787

Db    731 GCAAAAGAGATCCATCCCAGTGGCCAGAGAAAGAGGGAAAATGAAAGATTTTTACCTTGAATG 790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    788 ATTTTTGGCTTCATGCATCTTTCACTCACTCACCATATTGTAAACTGGTTCAGAAAATCTTAGCTC 847

Db    791 ATTTTTGGCTTCATGCATCTTTCACTCACTCACCATATTGTAAACTGGTTCAGAAAATCTTAGCTC 850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    848 GGATGTTGCTGAATCGACACTTCTTTAATGACCAGAATGAGATAAAATAGTATCTATAGTG 907

Db    851 GGATGTTGCTGAATGACACTTCCTTAATGACCAGAATAGAGAATTAGATAGTATCTATAGTG 910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    908 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGGAGGAAACACCTACTTTTGATAGGAGTGGAGA 967

Db    911 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGGAGAGAAACACCTACTTTTGATAGAGATGGAGA 970

Qy    968 AGGCATTTATGAAGGCTAGCTCTTGAACTCTTCAGTGACAGAAAAGCCATCTTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAAGGCTAGCTCTTGAACTCTTCAGTGCAGAAAAGCCAATCTTTACTTG 1030

Qy   1028 TATCAAATCAAAATGGAAAATATGTACACATTCTTCAGTATATGTTCCCATCGTGTTC 1087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAATCAAAATGGAAAATATGTACACATTCTTCAGTATATGTTCCCATCGTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAAGAAGGATTGGGAGTGTTTTTTCTTATTGGTT 1147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAAGAAGGATTGGGAGTGTTTTTTCTTATTGGTT 1150

Qy   1148 CTGGTTTGGCTGCCAACTCTGTACTCTTTAAAGTCACAACAAGATCGCTACACCGCGGGGTCTG 1207
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGCTGCCAACTCTGTACTCTTTAAAGTCACAACAAGATCGCTACACCGGGGGTCTG 1210

Qy   1208 CTCTCTTGATAAAATAAACAGCAAGTTTTATGTGATCTTAAATCAAAGGCTTGATTCAAGAACTG 1267
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTTATGTGATCTTAAATCAAAGGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGCACCAGATGTCTTCGCGTGAAAACATGAAAGCTCAGAGGAGCAGCACCCATCATTGG 1327
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCAGATGTCTTCGCGTGAAAACATGAAAGCTCAGAGGAGCAGCACCCATCATTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGGTTCAATAGAGATCCACTCTTTGAAGGAAGCGTGGTAACTTAGTTA 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAGCGTGGTAACTTAGTTA 1390

Qy   1388 GGGTTGGATGAAAAGCACAGAAGAGAACTTACGTCGGGGGTCCCACTCCAAATGATGAGACACTT 1447
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTTGGATGAAAAGCACAGAAGAGAACTTACGTCGGGGGTCCCACTCCAAATGATGAGACACTT 1450

Qy   1448 TGGATGAAGGAGTAGGACTTGTCGATTGCATTCAAAACTAGCAACTGAGCATATTCCAAGCCCTG 1507
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATTCCAAGCCCTG 1510

Qy   1508 CCAAGAAAGTACCAAGAGACTCCCTGCCACAGAGCAGCAAGAAACTTAGTCGGGGTGGGGGGT 1567
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGAGACTCCCTGCCACAGAGCAGCAAGAAACTTAGTCGGGGTGGGGGGT 1567

Qy   1568 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTCATGGGGGT 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTCATGGGGGT 1630

Qy   1628 GGGGGTATGGGAAACAGTTGG 1647
          |||||||||||||||||||||
Db   1631 GGGGGTATGGGAAACAGTTGG 1650
```

RESULT 14

AAH34834

```
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen encoding cDNA SEQ ID NO:1916.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma; chromosome 5; ss.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US026524.
XX
PR   29-SEP-1999;   99US-0157137P.
PR   03-NOV-1999;   99US-0163280P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   P-PSDB; AAG75429.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS   Claim 1; Page 3427-3428; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC   proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene therapy
CC   and vaccine production. N and P may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate P expression. For
CC   example, N and P may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of P by expressing inactive proteins or to
CC   supplement the patients own production of P. Additionally, N may be used
CC   to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the proteins. N and P
CC   can be used in the prevention, diagnosis and treatment of colorectal
CC   carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC   sequences used in the exemplification of the present invention. N.B.
CC   Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC   time of publication, meaning no sequences are present for SEQ ID NO:1027
CC   to 1052, 7921 and 7922
XX
SQ   Sequence 3008 BP; 944 A; 485 C; 638 G; 928 T; 0 U; 13 Other;

  Query Match             66.1%;  Score 1322.4;  DB 4;  Length 3008;
  Best Local Similarity   97.7%;  Pred. No. 9.5e-250;
  Matches 1373;  Conservative    9;  Mismatches   20;  Indels    4;  Gaps    4;

Qy         592 GGGCTTCGTGGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCT 651
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         221 GGGCTTCGTGGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCT 280

Qy         652 GAATATCCTATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGC 711
               |||||||||||||||||||||||||||||||||||||||||||||||||:||||||:|||
Db         281 GAATATCCTATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGSATTAGAMCGC 340

Qy         712 TGCTACTCTGTCTACTGCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAA 771
               ||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db         341 TGCTATTCTGTCTACTGCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGG-AATGATAAA 399

Qy         772 GATTTTACCTTGAATGATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTT 831
               |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db         400 GATTTTACCTTGAATGATTTTGGSTTCATGATCTTTCACTCACCATATTGTAAACTGGTT 459

Qy         832 CAGAAATCTCTAGCTCGGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAA 891
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         460 CAGAAATCTCTAGCTCGGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAA 519

Qy         892 AATAGTATCTATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTT 951
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         520 AATAGTATCTATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTT 579

Qy         952 GATAGAGATGTGGAGAAGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACA 1011
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         580 GATAGAGATGTGGAGAAGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACA 639

Qy        1012 AAGGCATCTTTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGT 1071
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         640 AAGGCATCTTTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGT 699

Qy        1072 TCCCTTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGA 1131
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         700 TCCCTTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGA 759

Qy        1132 GTGTTTTCTTATGGTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGAT 1191
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         760 GTGTTTTCTTATGGTTCTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGAT 819

Qy        1192 GCTACACCGGGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAAT-CAAG 1250
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db         820 GCTACACCGGGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCCAAG 879

Qy        1251 GCTTGATT-CAAGAACTGGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCT-CAGA 1308
               |||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||||
Db         880 GCTTGATTCCAAGAACTGGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCCAGA 939

Qy        1309 GAGGACACCCATCATTTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAA 1368
               ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db         940 GAGGACACCCATCATTKGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAA 999

Qy        1369 GGAACGTGGTACTTAGTTAGGGTGGATGAAAAGCACAGAAGAACTTACGCTCGGCGTCCC 1428
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1000 GGAACGTGGTACTTAGTTAGGGTGGATGAAAAGCACAGAAGAACTTACGCTCGGCGTCCC 1059

Qy        1429 ACTCCAAATGATGACACTTTGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACT 1488
               ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        1060 ACTCCAAATGRTGACACTTTGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACT 1119

Qy        1489 GAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCT 1548
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1120 GAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCT 1179

Qy        1549 GAAGCAGCTGTCATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAG 1608
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1180 GAAGCAGCTGTCATTAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAG 1239

Qy        1609 GGTGGGGTGGGCATGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGATATCTGGGGAT 1668
               |||:|||||:|||||||||||:|:||||||||||||||||||||||||||||||||||||
Db        1240 GGTKGGGTGSGCATGGGGTGGSGSTATGGGAACAGTTGGAGGAATGGGATATCTGGGGAT 1299

Qy        1669 AATTTTAAAGGATTACATGTTATGTAAATTTTTATGTGACTGACATGGAGCCTGGATGAC 1728
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1300 AATTTTAAAGGATTACATGTTATGTAAATTTTTATGTGACTGACATGGAGCCTGGATGAC 1359

Qy        1729 TATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTTGCTGACATGCTTCCTGTTGTGGTCT 1788
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1360 TATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTTGCTGACATGCTTCCTGTTGTGGTCT 1419

Qy        1789 GGCCAATGCCAAATGTACTCGAATGATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTG 1848
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1420 GGCCAATGCCAAATGTACTCGAATGATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTG 1479

Qy        1849 GCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGTGTACTTCTGT 1908
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1480 GCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGTGTACTTCTGT 1539

Qy        1909 CAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAAAATTTTTTAACATGTAAAAAAAAAA 1968
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
```

Db   1540 CAGAGAAAAGCAGAGGTACTAATCTCCAATTAAAAAATTTTTTAACATGTAAGAATTTTG 1599
            |||||||||||||||||||||||||||||||||||||||||||||
Qy   1969 AAAAAAAAAAAAAAAAAAAAAAAAA 1994

Db   1600 TACTTTGAACAACAAGATTACAGAAA 1625

RESULT 15
ACC62337
ID   ACC62337
XX
AC   ACC62337;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV411 encoding cDNA SEQ ID NO:203.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0326256P.
PR   17-SEP-2001; 2001US-0327781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323661P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0316663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;

PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54268.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 245-246; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1564 BP; 464 A; 304 C; 362 G; 434 T; 0 U; 0 Other;

Query Match   65.0%; Score 1302; DB 7; Length 1564;
Best Local Similarity   91.9%;
Matches 1438; Conservative   0; Mismatches   0; Indels 126; Gaps   1;

Qy   144 CATGCCTGGATCACTTCCTTTGAATCGCAGAAGCTTGCTGGCCAAAAGAATGTGGGAATTGT 203
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CATGCCTGGATCACTTCCTTTGAATCGCAGAAGCTTGCTGGCCAAAAGAATGTGGGAATTGT 60

Qy   204 TGCCCTTGAGATCTTATTTTTCCTTCTCCAATATGTTGATCAAGCACGAGTTGGAAAAATATGA 263
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGCCCTTGAGATCTTATTTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGA 120

Qy   264 TGGGTGGTAGATGCTGGGAAAGTATACCAATTGGCTTGGGCCCAGGCCAAGATGGGCTTCTGCAC 323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGGGTGTAGATGCTGGGAAAGTATACCAATTGGCTTGGGCCCAGGCCAAGATGGGCTTCTGCAC 180

Qy   324 AGATAGAAGAGATATTAACTCTCCTTTGCCATGACTGTGGTTCAGAATTCTTATGGAGAGAAA 383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGATAGAAGAGATATTAACTCTCCTTTGCCATGACTGTGGTTCAGAATTCTTATGGAGAGAAA 240

Qy   384 TAACCTTTCCTATGATTTGCATTGGGCGCGCTGGAAGTTGGAACAGAGACAATCATCGACAA 443
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TAACCTTTCCTATGATTTGCATTGGGCGCGCTGGAAGTTGGAACAGAGACAATCATCGACAA 300

Qy   444 ATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGAAATACAGATAT 503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGAAATACAGATAT 360

Qy   504 AGAAGGAAATCGACACAACTAATGCATGCTATGGAGGCCACCAGCTGCTTCTTCAATGCTGT 563
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGAAGGAAATCGACACAACTAAATGCATGCTATGGAGGCCACCAGCTGCTTCTTCAATGCTGT 420

Qy   564 TAACTCGGATTGAGTCCAGCTCTTGGGAT-------------------------------- 591
          ||||||||||||||||||||||||||||||
Db   361 AGAAGGAAATCGACACAACTAAATGCATGCTATGGAGGCCACCAGCTGCTTCTTCAATGCTGT 420

```
Qy   1578 TTAA 1581
          ||||
Db   1561 TTAA 1564

Search completed: June 24, 2004, 07:33:29
Job time : 783 secs


Db    421 TAACTGGATTGAGTCCAGCTCTTGGGGATGGACGGTATCCCTGGTAGTTGCAGGAGATAT 480
Qy    592 ---------------------------------------------------------- 591

Db    481 TGCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGGAGCAGAGTAGCTCTGCT 540
Qy    592 ------------------------------GGGCTTCGTGGGACACATATGCAACA 617
                                        ||||||||||||||||||||||||||
Db    541 AATTGGGGCCAAATGCTCCTTTAATTTTGAACGGAGGGCTTCGTGGGACACATATGCAACA 600

Qy    618 TGCCTATGATTTTTACAAGCCTGATATCGTCTATCTGAAATATCCTATAGTAGAGGAAAACT 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGCCTATGATTTTTACAAGCCTGATATCGTCTATCTGAAATATCCTATAGTAGAGGAAAACT 660

Qy    678 CTCCATACAGTGCTACCTCAGTGCATTAGAACCGCTGCTACTCTGTCTCTACTGCAAAAAGAT 737
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTCCATACAGTGCTACCTCAGTGCATTAGAACCGCTGCTACTCTGTCTCTACTGCAAAAAGAT 720

Qy    738 CCATGCCCAGTGGCAGAAAGAGGGAAAATGATAAAGATTTTACCTTGAATGATTTTTGGCTT 797
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCATGCCCAGTGGCAGAAAGAGGGAAAATGATAAAGATTTTACCTTGAATGATTTTTGGCTT 780

Qy    798 CATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTCCT 857
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTCCT 840

Qy    858 GAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTTGAAGC 917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTTGAAGC 900

Qy    918 CTTTGGGGATGTGTTAAATTTAGAGAGACACCTACTTTTTGATAGAGATGTGGAGAAGGCATTTAT 977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CTTTGGGGATGTGTTAAATTTAGAGAGACACCTACTTTTTGATAGAGATGTGGAGAAGGCATTTAT 960

Qy    978 GGAGGCTAGCCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTGTATCAAAATCA 1037
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GAAGGCTAGCCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTGTATCAAAATCA 1020

Qy   1038 AAATGGAAATATGTACACATCTTCAGTATATAGGTGGTTCCCTTGCATCGTTCTAGCACAGTA 1097
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAATGGAAATATGTACACATCTTCAGTATATAGGTGGTTCCCTTGCATCGTTCTAGCACAGTA 1080

Qy   1098 CTCACCTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTTTCTTATGGTTTCTGGTTTTGGC 1157
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCACCTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTTTCTTATGGTTTCTGGTTTTGGC 1140

Qy   1158 TGGCCACTCTGTACTCTCTTAAAGTCACACAAGAGTGCTACACCGGGGTCTGCCTCTTGATAA 1217
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TGGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCCTCTTGATAA 1200

Qy   1218 AATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGGTGTGGCACC 1277
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGGTGTGGCACC 1260

Qy   1278 AGATGTCTTCGCTGAAAACATGAAGAGCTCAGAGAGGACACCCATCATTTGGTCAACTATAT 1337
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGATGTCTTCGCTGAAAACATGAAGAGCTCAGAGAGGACACCCATCATTTGGTCAACTATAT 1320

Qy   1338 TCCCCAGGGTTCAATAGATTCACTCTTGAAGGAACGTGGTACTTAGTTAGGGTGGGGATGA 1397
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TCCCCAGGGTTCAATAGATTCACTCTTGAAGGAACGTGGTACTTAGTTAGGGTGGGGATGA 1380

Qy   1398 AAAGCACAGAAGAACTTAGCGTCGGGCGTCCACTCCACTCCAAATGATGACACTTTGGATGAAGG 1457
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAAGCACAGAAGAACTTAGCGTCGGGCGTCCACTCCAAATGATGACACTTTGGATGAAGG 1440

Qy   1458 AGTAGGACTTGTGCATTCAAAACATAGACACTAATTCCAAGCCCTGCCAAGAAAGT 1517
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 AGTAGGACTTGTGCATTCAAAACATAGACATAATTCCAAGCCCTGCCAAGAAAGT 1500

Qy   1518 ACCAAGAGACTCCCTGCCACAGCAGCAGCAGAGAACCTGAAGCAGCAGCTGTCATTAGTTAATGGGGAACA 1577
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 ACCAAGACTCCCTGCCACAGCAGCAGCAGAGAACCTGAAGCAGCAGCTGTCATTAGTTAATGGGGAACA 1560
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: June 24, 2004, 06:29:36 ; Search time 171 Seconds
(without alignments)
6497.144 Million cell updates/sec

Title: US-10-622-516-1
Perfect score: 2002
Sequence: 1 cgcctcccagcgactctcgg..........aaaaaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	2002	4	US-10-622-516-1	Sequence 1, Appli
2	2002	100.0	2002	4	US-09-819-993-1	Sequence 1, Appli
3	794.6	39.7	1824	1	US-10-193-295-1	Sequence 1, Appli
4	476	23.8	28001	4	US-08-305-505-1	Sequence 1, Appli
5	476	23.8	28001	4	US-09-819-993-3	Sequence 3, Appli
6	436.4	21.8	615	4	US-09-370-838-245	Sequence 245, App
7	410	20.5	500	4	US-09-370-838-128	Sequence 128, App
8	401	20.0	506	4	US-09-401-064-187	Sequence 187, App
9	224	11.2	448	4	US-09-833-381-1645	Sequence 1645, Ap
10	210.4	10.5	472	4	US-09-833-381-1648	Sequence 1648, Ap
11	135.8	6.8	635	4	US-09-998-416-730	Sequence 730, App
12	110	5.5	307	4	US-09-833-381-744	Sequence 744, App
13	103.6	5.2	4775	3	US-09-306-595C-1	Sequence 1, Appli
14	103.6	5.2	4775	3	US-09-925-388-1	Sequence 1, Appli
15	86.6	4.3	305	4	US-09-313-294A-4313	Sequence 4313, Ap
16	74.6	3.7	385	4	US-09-833-381-1646	Sequence 1646, Ap
17	71.2	3.5	288	4	US-09-313-294A-5701	Sequence 5701, Ap
18	70.4	3.5	593	4	US-09-904-615-59	Sequence 59, Appl
19	70	3.5	293	4	US-09-313-294A-7134	Sequence 7134, Ap
20	67.2	3.4	232	4	US-09-833-381-745	Sequence 745, App
21	64.4	3.2	664	4	US-09-904-615-66	Sequence 66, Appl
22	64	3.2	282	4	US-09-621-976-18648	Sequence 18648, A
23	63.4	3.2	2550	6	US-525287-23	Patent No. 5258287
24	62	3.1	6409	4	US-09-967-908A-1	Sequence 1, Appl
25	59.8	3.0	1474	1	US-08-821-994-64	Sequence 64, Appl
26	59.6	3.0	194	4	US-09-621-976-9596	Sequence 9596, Ap
27	59.6	3.0	3581	2	US-08-738-349-1	Sequence 1, Appli
28	59.2	3.0	1582	3	US-08-545-196B-10	Sequence 10, Appl
29	59.2	3.0	1582	3	US-08-545-196B-12	Sequence 12, Appl
30	59.2	3.0	2186	4	US-09-360-545-66	Sequence 66, Appl
31	59	3.0	1049	4	US-09-800-729-67	Sequence 67, Appl
32	58.8	2.9	1641	1	US-08-300-903A-8	Sequence 8, Appli
33	58.8	2.9	1641	4	US-09-182-816-22	Sequence 22, Appli
34	58.8	2.9	1736	3	US-09-182-816-24	Sequence 24, Appli
35	58.8	2.9	1736	3	US-09-471-528-22	Sequence 22, Appli
36	58.8	2.9	1736	3	US-09-471-528-24	Sequence 24, Appli
37	58.8	2.9	1736	3	US-09-634-530-22	Sequence 22, Appli
38	58.8	2.9	1736	3	US-09-634-530-24	Sequence 24, Appli
39	58.8	2.9	1736	3	US-09-363-708-3	Sequence 3, Appli
40	58.8	2.9	2246	4	US-09-083-587-3	Sequence 3, Appli
41	58.8	2.9	2246	4	US-09-248-335-67	Sequence 67, Appli
42	58.6	2.9	1074	3	US-09-601-537-10	Sequence 10, Appl
43	58.2	2.9	441	4	US-09-247-373B-33	Sequence 33, Appl
44	58.2	2.9	1117	3	US-09-601-537-9	Sequence 9, Appl
45	58.2	2.9	4121	4	US-09-601-537-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-819-993-1
Sequence 1, Application US/09819993
Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-09-819-993-1

Query Match 100.0%; Score 2002; DB 4; Length 2002;
Best Local Similarity 100.0%;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 CGCCTCCCAGCGACTCTCGGCAGTCCCGGAGTCGGGTTGGCGGGCTATAAAGCTGGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CGCCTCCCAGCGACTCTCGGCAGTCCCGGAGTCGGGTTGGCGGGCTATAAAGCTGGT 60

Qy    61 AGCCAAAGGGGAGGGGGCGGGCCCGGGGACTGTCCTTCTTGTGGGCTCACTCCCTTTCCTCTTCTGTCGCG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AGCCGAAGGGGAGGGGGCGGGCCCGGGGACTGTCCTTCTTCGTGTGGCTCACTCCCTTTCCTCTTCTGTCCG 120

Qy   121 CTCGGTCACGCTTGCTCTTTCACCATGCCTGGGATCACTTCCTTTGAAATGCAGAAGAAGCTTGC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CTCGGTCACGCTTGCTCTTTCACCATGCCTGGGATCACTTCCTTTGAAATGCAGAAGAAGCTTGC 180

Qy   181 TGGCCAAAACGATGTCGGGAATTGTTGCCCCTTGAGATCTATTTTCCTTCANTATATGTTGAT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGGCCAAAAGATGTGGGGAATTGTTGCCCCTTGAGATCTATTTTCCTTCAATATATGTTGAT 240

Qy   241 CAAGCAGAGTTGGAAAAAATATGATGGTGTAGATCGTGTAGAATGTGTAGTGTAGCTTGGGC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAAGCAGAGTTGGAAAAAATATGATGGTGTAGATCGTGTAGATCGTGTAGTGTAGCTTGGGC 300

Qy   301 CAGGGCCAAGATGGGGCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTTGCATGACTGTG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGGCCAAGATGGGGCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTTGCATGACTGTG 360

Qy   361 GTTCAGAATCTTATGGAGAGAAAATAACCTTTCCTATGATTGATTGCATTGGGCGGCTGGAAGTT 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTTCAGAATCTTATGGAGAGAAAATAACCTTTCCTATGATTGATTGCATTGGGCGGCTGGAAGTT 420
```

```
Qy    361 GTTCAGAATCTTATGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTT 420
Db    361 ......................................................................

Qy    421 GGAACAGAGACAAATCATCGACAAATCAAAGTCTGTGAAGACTAATTGATTGAATGCAGCTGTTT 480
Db    421 GGAACAGAGACAAATCATCGACAAATCAAAGTCTGTGAAGACTAATTGATTGAATGCAGCTGTTT 480

Qy    481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGACAACAACTAATGCATGCTATAGGAGGC 540
Db    481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGACAACAACTAATGCATGCTATAGGAGGC 540

Qy    541 ACAGCTGCTGTCTTCAAATGCTTAACTGGGATTGAGTCCAGCTCTTGGGAATGGGCTTCGT 600
Db    541 ACAGCTGCTGTCTTCAAATGCTTAACTGGGATTGAGTCCAGCTCTTGGGAATGGGCTTCGT 600

Qy    601 GGGACACATATGCAACATGCCTATGATTTTACAGAGCCTAGTACTCTGAATATCCT 660
Db    601 GGGACACATATGCAACATGCCTATGATTTTACAGAGCCTAGTACTCTGAATATCCT 660

Qy    661 ATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCCCTGCTACTCT 720
Db    661 ATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCCCTGCTACTCT 720

Qy    721 GTCTACTGCAAAAAGATCCATGCCCAGTGGCGAGAAGAAAATGATAAAGAGATTTTACC 780
Db    721 GTCTACTGCAAAAAGATCCATGCCCAGTGGCGAGAAGAAAATGATAAAGAGATTTTACC 780

Qy    781 TTGAATGATTTGGCTTCATGATCTTTCACTCACCACCATATTGTAAACTGGTTCAGAAATCT 840
Db    781 TTGAATGATTTGGCTTCATGATCTTTCACTCACCACCATATTGTAAACTGGTTCAGAAATCT 840

Qy    841 CTAGCTCGGATGTTGCTGAATGACTTCCTTAATCACCAGAATAGAGATTAGTAGTATC 900
Db    841 CTAGCTCGGATGTTGCTGAATGACTTCCTTAATCACCAGAATAGAGATTAGTAGTATC 900

Qy    901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGGAGCACCTACTTTGATAGAGAT 960
Db    901 TATAGTGGCCTGGAAGCCTTTGGGGATGTTAAATTAGGAGCACCTACTTTGATAGAGAT 960

Qy    961 GTGGGAGAGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCGTCGAGAAAACAAAGGCATCT 1020
Db    961 GTGGGAGAGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCGTCGAGAAAACAAAGGCATCT 1020

Qy   1021 TTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080
Db   1021 TTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080

Qy   1081 TCTGTTCTTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTCT 1140
Db   1081 TCTGTTCTTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTCT 1140

Qy   1141 TATGGTTCTGGGTTTGGGCTGCCACTCTGTACTCTTAAAGTCACACAAGAGATGCTACACCG 1200
Db   1141 TATGGTTCTGGGTTTGGGCTGCCACTCTGTACTCTTAAAGTCACACAAGAGATGCTACACCG 1200

Qy   1201 GGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGACTTAAATCAAGGCTTGATTTCA 1260
Db   1201 GGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGACTTAAATCAAGGCTTGATTTCA 1260

Qy   1261 AGAACTGGTGTGGGCACCAGATGTCTTCGCTGAAAACATGAAGCTTCAGAGAGGACACCCAT 1320
Db   1261 AGAACTGGTGTGGGCACCAGATGTCTTCGCTGAAAACATGAAGCTTCAGAGAGGACACCCAT 1320

Qy   1321 CATTTGGTCAACTATATTCCCAGGGTTCAATAGAATTCACTCTTTGAAGGAACGTGGTGTAC 1380
Db   1321 CATTTGGTCAACTATATTCCCAGGGTTCAATAGAATTCACTCTTTGAAGGAACGTGGTGTAC 1380

Qy   1381 TTAGTTAGGGTGGATGAAAAGCACAGAGAGAACTTACGGCTTCGGGGTCCCACTCCAAATGAT 1440
Db   1381 TTAGTTAGGGTGGATGAAAAGCACAGAGAGAACTTACGGCTTCGGGGTCCCACTCCAAATGAT 1440

Qy   1441 GACACTTTGGATGAAGGAGTAGGAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCA 1500
Db   1441 GACACTTTGGATGAAGGAGTAGGAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCA 1500
```

Query Match   100.0%; Score 2002; DB 4; Length 2002;
Best Local Similarity   100.0%;
Matches 2002; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy      1 CGCCTCCCAGCGACTCTCGGCCAGTGCGCGGAGTCGGGGTGCGGGGCTATAAAGTCTGGT 60
Db      1 CGCCTCCCAGCGACTCTCGGCCAGTGCGCGGAGTCGGGGTGCGGGGCTATAAAGTCTGGT 60

Qy     61 AGCGAAGGGGAGGCGCCCCGGGACTGTCCTTTCGCTGGCTCACTCCCTTTCCTCTGCCG 120
Db     61 AGCGAAGGGGAGGCGCCCCGGGACTGTCCTTTCGCTGGCTCACTCCCTTTCCTCTGCCG 120

Qy    121 CTCGGTCACGCGCTTGCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGC 180
```

Qy / Db sequence alignment, positions 121 through 2002, displayed as paired rows with match bars (|||||...). Example paired rows:

Db 121 CTCGGTCACGCTTGCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGC 180
Qy 121 CTCGGTCACGCTTGCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGC 180

... (alignment continues through positions ~180 to 2002) ...

Db 1981 AAAAAAAAAAAAAAAAAAA 2002
Qy 1981 AAAAAAAAAAAAAAAAAAA 2002

RESULT 3
US-08-305-505-1
; Sequence 1, Application US/08305505
; Patent No. 5668001
GENERAL INFORMATION:
; APPLICANT: Miziorko, Henry M.
; TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
; TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
; TITLE OF INVENTION: STABILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:

```
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/305,505
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/08/072,040
;     FILING DATE: 02 JUNE 1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Baker, Jean C.
;     REGISTRATION NUMBER: 35,433
;     REFERENCE/DOCKET NUMBER: 65-053-9083-9
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (414) 277-5709
;     TELEFAX: (414) 277-5591
;   INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1824 base pairs
;       TYPE: nucleic acid
;       STRANDEDNESS: double
;       TOPOLOGY: linear
;     MOLECULE TYPE: cDNA
US-08-305-505-1

Query Match    39.7%;  Score 794.6;  DB 1;  Length 1824;
Best Local Similarity   72.3%;  Pred. No. 4e-185;
Matches 1139;  Conservative 0;  Mismatches 304;  Indels 132;  Gaps 3;
```

```
Qy    144 CATGCCTGGATCACTTCCTTTGAAATGCAGAGAAGCTTGCTGGCCAAAAAAGATGTGGGGAATTGT 203
          |||| ||||| || ||| ||| ||||| ||||||||  |||||||||||       |||||| ||||
Db      1 CATGGCTGGGTCGTCTCTTCTCCAGTGAACACTGAATCGAAGCTTGCTGGCCAAAGATGTGGGTTATTGT 60

Qy    204 TGCACTGGAAATCTATTTTCCCTCTCAGTGATGTCGACCAGCTGAGCTGGAGAAGGTATGA 263
          |||| || ||| || ||| ||| || |||||  |||||||||  |||||||||||  |||
Db     61 TGCACTGGAAATCTATTTTCCCTCTCAGTGATGTCGACCAGCTGAGCTGGAGAAGGTATGA 120

Qy    264 TGGGTGTAGAGTGCTGGAAAGTATACCATTGGCTTGGGCCCAAGAGATGGCTTCTGCAC 323
          ||| || ||| |||| ||| || ||||| |||||||||||| ||||||||||| |||
Db    121 CGGTGTGGATGCAGGCAAAATACACCATTGGGTTAGGCCCAAGAGATGGCTTCTGCTC 180

Qy    324 AGATAGAGAAGAGATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGCGAGAGAAA 383
          |||||||  ||||  ||||| || || ||||||||||||| || |||||||||||| |||
Db    181 TGACCGAGAGGGAATATCAAATTCCCTCGTTGACTGTCGGTTCAGAGAGCTTATGGAGGAA 240

Qy    384 TAACCTTTCCTATGATTCCATTGGGCGGCGCGGCGGAAGGTTGGAACAGAGACAAATCATCGACAA 443
          ||||  || |||| | |||||| |||    ||||| || ||||||||||| |||||||||| ||||
Db    241 CAGCCTTTCCTATGACTGTTGTTGGGAGGAGTTGGAACCGGAGCAATAATTGATAA 300

Qy    444 ATCAAAGTCGTGTGAAGACTAATTTGATGCAGCGCTGTTTTGAAGAGTTCTGGGAATTACAGATAT 503
          |||||| || ||||||| ||||||| || |||||| || ||||||| |||||| ||| |||||||
Db    301 ATCAAAATCGGTGAAGACTGTCCTGATCGCAGCTATTTTGAAGAATCTGGTAATACAGATGT 360

Qy    504 AGAGGAAATCGACAACAACTACTAGTGTCCTGTAACGGTGCTAGTGGAGGCACACGCTGTGCTT 563
          |||| ||| || ||||| ||| ||| || |||||||| |||||||||||||| || |||||
Db    361 AGAAGGAAATTGACAACAACCAATGCGTCGTATGGAGGCACACTGCTGTCTGCTT 420

Qy    564 TAACTCGATTGAGTCCAGCTCTTGGGATGGGGC---------------------------- 595
          |||||| ||| ||| || ||||||  ||  |||
Db    421 TAACTGGATTGAGTCCAGTTCTCTGGGATGGGGACGCTATGCCACTTGTTGTTGCTGGGACAT 480

Qy    596 ---------------------------------------------------------------- 595
Db    481 TGCTGTCTGTATGCCACTGGGAGCTGGGAGCGCAACAGGTGAGCTGGTGCTGTTGCTATGCT 540

Qy    596 ---------------------------------------------------------------- 595
Db    541 AGTTGGGGTCAAAATGCTCCTTTAATTTTTGAGAGGAATTGCGTGGGAACCCACATGCAGCA 600

Qy    618 TGCCTATGATTTTTACAGCCTGATAATGCTCCTCTAGTAATATCCTATAGTAGATGATGGAAAACT 677
          |||| ||| || || ||  ||||  |   ||   |||||||
Db    601 TGCTTATGACTTCTATAAACCAGAATATGGTTTCTGAATAATCCTGTTAGTTGACTGGCAAACT 660
```

```
Db    601 TGCTTATGACTTCTATAAACCAGAATATGGTTTCTGAATAATCCTGTTAGTTGACTGGCAAACT 660
          |||| ||| || || ||  ||||  | |||| |||| |||||||| ||| ||| |||||| ||||
Qy    678 CTCCATACAGTGTCACCTTCAGTGCAATTGAGACCGCTGCCTACTCTGTCTGCAAAGAT 737

Db    661 ATCTATACAGTGCTACCTCACTTCAGTGCATTAGGACCGGTCATTAGAGTGTTTATCGAATAAAAT 720
          | ||||||||| ||||| || ||||||||| || ||| || ||||| ||||| |||||||||||
Qy    738 CCATGCCCAGTGGCGAAAGAGGGGAAATGAATAAAGATTTAGCCTTGGAATTGATTTTTGCTT 797

Db    721 CCATGCCCAGTGGCGAAAAGAGGGGACAGACAGAGTTTCACCTTGAACTTTTGGATT 780
          |||||||||||||||||  |||||||| |||| |||||||||| |||||||  ||||
Qy    798 CATGATCTTTCACTCACCAATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCT 857

Db    781 CATGATCTTTCATTCTCCTACTGTTAAACTGGTTCAGAAATCTCTAGCTCGGAGAGAGCT 840
          ||||| ||||||| || ||| ||| |||||||||||||||||||||||| |  |||||
Qy    858 GAATGAGACTTCCTTAATGACCAGAATAGAGATTA------AAAATAGTTATCTATAGTGACACA 914

Db    841 GAATGACTTTCTAGCTGACAAATGCAGAAAACAGCAAATGGTTGTTTCAGTGCTGCGATTA 900
          ||||||  || ||  ||||  || ||||  || || |||||| |||  |||| ||| || 
Qy    915 AGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATGATGAGAGTGGAGAAGGCATT 974

Db    901 AGCTTTCAGGGAGTGTAAAGCTTGAAGATACAATATTTTGATAGGATGGAAAAAGCTTT 960
          ||| ||| ||||  |||||  || ||||  |||||||| |||| || |  ||||| || 
Qy    975 TATGAGAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTGTATCAAA 1034

Db    961 TATGAAAGCTAGTGCAGAGCTCTTCAATCAGAAAACCAAAGCTTCCTTACTTGTCTCCAA 1020
          |||||  | ||| ||||| ||||||  |||||||||| |||| || ||||||||| || ||
Qy   1035 TCAAAATGGAAATATGTACACAATCTTCAGTATATTAGGTTCCCTTGCATCTGTTCTAGCA 1094

Db   1021 TCAGAATGGAAACATGTACACGCCCTTCAGTGCTACGGTTTGCCTTGCTTTCTCTTTAGCCA 1080
          |||| |||||||| |||||||| | ||||||| || ||||| ||||||  ||| |||||| |
Qy   1095 GTACTCACCTCAGCAATTAGCACGGGAAGAGAAGAATCAGTCGTTTTTGCTTTGTTTGTTT 1154

Db   1081 GTACTTCAGAGCGACAGCTTCAGGGAAGAAGTCAGTTTCTCATAGGCCTCTGGGTTT 1140
          |||||| || || || |||  |||||  ||| ||| |||| || || ||| ||||||
Qy   1155 GGCTGGCCACTCGTACTCTCTTAAAGTCACACACAAGAGTCTACACGCGGGTTTCTGCTTGA 1214

Db   1141 TGCTGGCTACGGTGTAATTCCATCAGAGTTACACAGAGATGCCACTCCTGGTTTCTCGCTGA 1200
          ||||||| || || |||||  | ||||||||||| ||||||||||| ||| ||||| ||||
Qy   1215 TAAAATAACAGCAAAGTTTAATGGATCTTAAATCAAGGCCTTGAATTGGGTGGGC 1274

Db   1201 CAAAAATAACTGCTAGCCTTTCTGATCTTAAAGCAACAACTGTAACTCACGAAAATGCATTGC 1260
          ||||| ||| ||| | ||||| ||||||||| || || |||| || ||| | |||| |||||
Qy   1275 ACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAAGAAGCACACCCATCATTTTGTTCACTA 1334

Db   1261 ACCTGATGTCTTTCGCTGAAAAACATGAAGATTAGACAGAGAGCACAATCACTTGGCCAACTA 1320
          ||| |||||||| ||||||||||||||||||  |||||||||||| |||||| |||| ||||
Qy   1335 TATTCCCCAGGGTTCAATAGATCTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGATGGA 1394

Db   1321 TATTCCACACTGTTCAGTGAGAAGATCTCTTTTGTGCCTGTCCTGTGTGGA 1380
          ||||||| || ||||| ||| |||||||| ||||||   ||| |||||||
Qy   1395 TGAAAAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATCATGAGACACTTTGGGATGA 1454

Db   1381 TGAAAAACACAGGAGAACAATATGCACGACAGCCCAGTTATGGGTGATGGAGGC 1440
          |||||| |||| ||||| || || ||  | |||||||| || ||  | |||| ||
Qy   1455 AGGAGTAGGACTTGTGCATTCAAACATAGCAACCTGAGCATATTCCAAGCCTCCCAAGAA 1514

Db   1441 AGGAGTTGAAGTTGTCCACCAGGAGTTCAGCAGAAGTACCGCCTGGTTCCACTGGAAGAA 1500
          ||||| | || ||| |||  ||||||||| |||||| ||  | | |  |||  |||||||
Qy   1515 AGTACCAAGACTCCCTGCCCACAGCCAG---CAGAACCTGAAGCAGCTGGTCCATTCATTAGTTAATGG 1571

Db   1501 AGTGCCAAGAATCCCTGCAAACAACACAGAAGCGTTACTGTTGCCATTCCAATGG 1560
          ||| |||||| |||||||  |||  ||||||| | ||||| | ||||| | ||||
Qy   1572 GGAACATTAAGATAC 1586

Db   1561 GGTGCATTAAGATAC 1575
          || ||||||||||||
```

; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

Query Match   23.8%;   Score 476;   DB 4;   Length 28001;
Best Local Similarity   95.2%;
Matches   491;   Conservative   0;   Mismatches   25;   Indels   0;   Gaps   0;

```
Qy   1479 CATAGCAACTGAGCATATTCCAAGCCCTGCCACAAGAAGTACCAAGACTCCCTGCCACAGC 1538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24597 CCTTTCATTTAGCATATTCCAAGCCCTGCCACAAGAAGTACCAAGACTCCCTGCCACAGC 24656
Qy   1539 AGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGGAACATTAAGATACTCTGTGAGGTGC 1598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24657 AGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGGAACATTAAGATACTCTGTGAGGTGC 24716
Qy   1599 AAGACTTCAGGGTGGGGTGGGGTGGGGTGGGGTGGGGGTGGGGAATGGGAT 1658
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24717 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGGGAACAGTTGGAGGAATGGGAT 24776
Qy   1659 ATCTGGGGATAATTTTAAAGGATTACATGTTATGTTATGTGACTGACATGGAG 1718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24777 ATCTGGGGATAATTTTAAAGGATTACATGTTATGTTATGTGACTGACATGGAG 24836
Qy   1719 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTGCTGACATGCTTCCT 1778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24837 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTGCTGACATGCTTCCT 24896
Qy   1779 GTTGTGGTCTCGGCCAAATGCCAAATGTACTTAAGGGGCTCTTGTAAAAACTTCA 1838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24897 GTTGTGGTCTCGGCCAAATGCCAAATGTACTTAAGGGGCTCTTGTAAAAACTTCA 24956
Qy   1839 TACCTCTTTGGCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGT 1898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24957 TACCTCTTTGGCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGT 25016
Qy   1899 GTACTTCTGTCAGAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAAAAATTTTTTAACATGT 1958
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25017 GTACTTCTGTCAGAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAAAAATTTTTTAACATGT 25076
Qy   1959 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1994
          |||||||||||||||||||||||||||||||||||||||||||||
Db  25077 AAGAATTTTGTACTTTGAACAACAAGATTACAGAAA 25112
```

RESULT 5
US-10-193-295-3
; Sequence 3, Application US/10193295
; Patent No. 662608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29

; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(28001)
; OTHER INFORMATION: n = A,T,C or G
US-10-193-295-3

Query Match   23.8%;   Score 476;   DB 4;   Length 28001;
Best Local Similarity   95.2%;
Matches   491;   Conservative   0;   Mismatches   25;   Indels   0;   Gaps   0;

```
Qy   1479 CATAGCAACTGAGCATATTCCAAGCCCTGCCACAAGAAGTACCAAGACTCCCTGCCACAGC 1538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24597 CCTTTCATTTAGCATATTCCAAGCCCTGCCACAAGAAGTACCAAGACTCCCTGCCACAGC 24656
Qy   1539 AGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGGAACATTAAGATACTCTGTGAGGTGC 1598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24657 AGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGGAACATTAAGATACTCTGTGAGGTGC 24716
Qy   1599 AAGACTTCAGGGTGGGGTGGGGTGGGGTGGGGGTGGGGAATGGGAT 1658
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  24717 AAGACTTCAGGGTGGGGTGGGCATGGGGTGGGGTATGGGGGAACAGTTGGAGGAATGGGAT 24776
Qy   1659 ATCTGGGGATAATTTTAAAGGATTACATGTTATGTTATGTGACTGACATGGAG 1718
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24777 ATCTGGGGATAATTTTAAAGGATTACATGTTATGTTATGTGACTGACATGGAG 24836
Qy   1719 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTGCTGACATGCTTCCT 1778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24837 CCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTTGCTCTATTGCTGACATGCTTCCT 24896
Qy   1779 GTTGTGGTCTCGGCCAAATGCCAAATGTACTTAAGGGGCTCTTGTAAAAACTTCA 1838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24897 GTTGTGGTCTCGGCCAAATGCCAAATGTACTTAAGGGGCTCTTGTAAAAACTTCA 24956
Qy   1839 TACCTCTTTGGCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGT 1898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24957 TACCTCTTTGGCCATTTGTATGCATGATGTTTGGTTTTTAAACATGGTATAATGAATTGT 25016
Qy   1899 GTACTTCTGTCAGAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAAAAATTTTTTAACATGT 1958
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25017 GTACTTCTGTCAGAGAAGAAAGCAGAGGTACTAATCTCCAATTAAAAAAATTTTTTAACATGT 25076
Qy   1959 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1994
          |||||||||||||||||||||||||||||||||||||||||||||
Db  25077 AAGAATTTTGTACTTTGAACAACAAGATTACAGAAA 25112
```

RESULT 6
US-09-370-838-245
; Sequence 245, Application US/0937083B
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 245
; LENGTH: 615

```
;
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (105)
;  OTHER INFORMATION: n=A,T,C or G
US-09-370-838-245

Query Match    21.8%;  Score 436.4; DB 4; Length 615;
Best Local Similarity  99.1%; Pred. No. 1.4e-97;
Matches  459; Conservative  0; Mismatches  2; Indels  2; Gaps  2;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
Db    92 GCTCTTTCACCAT-NCTGGATCACACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 150

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db   151 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 210

Qy   254 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 313
Db   211 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 270

Qy   314 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
Db   271 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 330

Qy   374 TGGAGAGGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGCTGGAAGTTGGAAACAGAGACAA 433
Db   331 TGGAGAGAAAATAACCTTTCCTATGATTGCATTGGGCGGCGGCTGGAAGTTGGAAACAGAGACAA 390

Qy   434 TCATCGACAAATCAAAGTCTCTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGGA 493
Db   391 TCATCGACAAATCAAAGTCTCTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGGA 450

Qy   494 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGAGGCACA 543
Db   451 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGAGGCACA 500

Qy   554 TCAATGC-TGTTAACTGGATTGAGTCCAGCTCTTGGGGATGGGC 595
Db   511 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGGATGGGAC 553
```

```
RESULT 7
US-09-370-838-128
; Sequence 128, Application US/09370838
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 128
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-128

Query Match    20.5%;  Score 410; DB 4; Length 500;
Best Local Similarity 100.0%; Pred. No. 3.9e-91;
Matches  410; Conservative  0; Mismatches  0; Indels  0; Gaps  0;
```

```
RESULT 8
US-09-401-064-187
; Sequence 187, Application US/09401064
; Patent No. 6623923
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 187
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-401-064-187

Query Match    20.0%;  Score 401; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 6.2e-89;
Matches  401; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
Db    91 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 150

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db   151 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 210

Qy   254 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 313
Db   211 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 270

Qy   314 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
Db   271 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 330

Qy   374 TGGAGAGGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGCTGGAAGTTGGAAACAGAGACAA 433
Db   331 TGGAGAGGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGCTGGAAGTTGGAAACAGAGACAA 390

Qy   434 TCATCGACAAATCAAAGTCTCTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGGA 493
Db   391 TCATCGACAAATCAAAGTCTCTGAAGACTAATTTGATCGAAGCTGTTTGAAGAGTCTGGGA 450

Qy   494 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGAGGCACA 543
Db   451 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGAGGCACA 500
```

```
Query Match    20.0%;  Score 401; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 6.2e-89;
Matches  401; Conservative  0; Mismatches  0; Indels  0; Gaps  0;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
Db   106 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 165

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
Db   166 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 225

Qy   254 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 313
Db   226 AAAAAATATGAGATGGTGTAGAATCGCTGGAAAGTATACCATTGGCTTGGGCCCAAGATGG 285

Qy   314 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
Db   286 GCTTCTCGCACAGATAGAGAAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 345
```

```
Qy   374 TGGAGGAGAAATAACCTTTCCTATGATTGCATTGGGGCGGCCTGGGCCTGCGGAAGTTGGAACAGAGAGACAA 433
         ||||| |||||||||||||||||||||||||||| |||||||||||||  |||||||||||||||| ||||||
Db   346 TGGAGAGAAAATAACCTTTCCTATGATTGCATTGGGCGGCCGGCTGCGGAAGTTGGAACAGAGACA 405

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTGACTGCAGCTGTTTGAAGAGTTCTGGGA 493
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   406 TCATCGACAAATCAAAGTCTGTGAAGACTAATTGACTGCAGCTGTTTGAAGAGTTCTGGGA 465

Qy   494 ATACAGATATAGAAGGAATCGACACACTAATGCATGCTAT 534
         ||||||||| |||||||||||||||||||||||||||||
Db   466 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTAT 506
```

RESULT 9
US-09-833-381-1645
Sequence 1645, Application US/09833381
Patent No. 6672186

GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1645
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(448)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1645

Query Match 11.2%; Score 224; DB 4; Length 448;
Best Local Similarity 82.1%; Pred. No. 1.4e-45;
Matches 262; Conservative 0; Mismatches 54; Indels 3; Gaps 2;

```
Qy  1142 ATGGTTCTTGGTTTGGCTGCCCACTCTGTACTCTCTTAAAGTCACACAGATCGTACACCGG 1201
         |||||||||||| |||  || ||||||||||  || ||| |||||||||||||||||| |||
Db     8 ATGGTTCTGGGTTGTGGCTCGCCACTCTCTNTANTCTCTTANATTCACACNNNAATGCTNCACCGG 67

Qy  1202 GGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAA 1261
         ||||||||||||||| |||||||||||  || || ||||||||||||||||||||||||||
Db    68 GGTCTGCTCTTGATACAATNNNTATCAANTTNATGTGATCTTAAATCAAGGCTTGATTCGA 127

Qy  1262 GAACTGGTGTGGCACCAGATGTGTTCTTCCGCTGAAAACATGAGAGGAGGCACCCCATC 1321
         ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   128 TAACTGGTGTGGCAACATATGTCTTCNCGNGANAANNNTGANACNTCATATANGACACCCATC 187

Qy  1322 ATTTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTTGAAGGAACGTGGTTACT 1381
         ||||||||  ||  || |||| |||||||  ||| |||||| | || ||||||  ||||| ||
Db   188 ATTTGGTCNACTNTATTCCCAATGGTTCNAATTCACTCTTTTGAAAGGAACNTGGGTNCT 247

Qy  1382 TAGTTA-GGGTGGATGAAAAGCACAGAAGAACTTACGCTTCGGGGCGTCCC--ACTCCAAATG 1438
         ||||||  ||||||||||| |||||||| |||||| |||||||||||||||   ||||||||||
Db   248 TACTTNTTGGTTGGATGATTAAGCNTANAAGAAAATTNCGCTCGGGAGTCCCACTCCCNAATN 307

Qy  1439 ATGACACTTTGGATGAAGG 1457
         |||||||||||||||||||
Db   308 ATGACACTTTGGATCAAAG 326
```

RESULT 10
US-09-833-381-1648
Sequence 1648, Application US/09833381
Patent No. 6672186

GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1648
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(472)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1648

Query Match 10.5%; Score 210.4; DB 4; Length 472;
Best Local Similarity 70.1%; Pred. No. 3e-42;
Matches 333; Conservative 0; Mismatches 129; Indels 13; Gaps 4;

```
Qy  1270 GTGGGCACCAGATGTCTTCCGCTGAAAACATGAA-GCTCAGAGGAGGAGCACCCATCATTTGGT 1328
         |||||||||||||||||||||||||||||||||  |||||||| || ||||| |||| ||||||
Db     8 GTGGCACCCGGATGTCTTCTGCTGAAAAACATGAATGCTCAGAGAGAGGACACATCACTTAGC 67

Qy  1329 CAACTATATTCCCCAGGGTCAATAGATTCACTCTTTGAAGGAAGATGGGTGGTTACTTAGTTAG 1388
         ||||||||||||||||    |  |  |||||||||||| ||||| ||||   |||| |||||||
Db    68 CAACTATATTCCCCA-TGTTCAATANACTCACTCTTTGAAGGAAGCTGGGTATCTGGTCAG 126

Qy  1389 GGTGGATGAAAAGACACAGAAGAACTTACGCTTCGCGGGGTCCCACTCCAAATAATGACACTTT 1448
         |||||||||||| |||||||| ||||| ||| || |||||| |||| |||  |  |||||| ||
Db   127 AGTGGATGAAAACACAGAAGGACTTACGGCCCGGCCCCTTCACAAATGACCACAGTTT 186

Qy  1449 GGATCAAGGAGTAGGAGACTTGTCATTCAAACATAGCAACTGAGCATAATTCCAAGCCCTGC 1508
         |||||||| |||||||| ||||||||||||||| |||||| ||||  || |||||||||||||
Db   187 GGATGAAGGAAATGGNNCTCGTGCATAGTAACACACACAGAGCAATATTCCAAGCCCTGC 246

Qy  1689 T--ATGTAAATTTTTATGTGACTGACATGGAGCCTGGATGAGACTATCGTGTTACTTG 1741
         |  ||| ||||||| ||||||||||||||||| |||| | |||| |||||||||||||
Db   418 TTGCTTAAAAATGNNATGGAACTGACACAGANCCCAGAAAGCTINTTGNGTTTTG 472
```

RESULT 11
US-08-998-416-730
Sequence 730, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road

CITY: Research Triangle Park
; STATE: No. 62392664th Carolina
; COUNTRY: USA
; ZIP: 27709
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
; NAME: Meige, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/S-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8689
; TELEFAX: 919-541-8587
INFORMATION FOR SEQ ID NO: 730:
SEQUENCE CHARACTERISTICS:
; LENGTH: 635 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
; ORGANISM: PAG1476UP

US-08-998-416-730

Best Local Similarity 62.6%;   Pred. No. 6.4e-24;   Indels 132;   Gaps 1;
Query Match   6.8%;   Score 135.8;   DB 3;   Length 635;
Matches 231;   Conservative 0;   Mismatches 132;   Indels 6;   Gaps 1;

Qy   229 CAATATGTTGATCAAGCACAGAGTTGGAAAAATATGATTGGTGTGTAGAATGCTGGAAGTATACC 288
          || |||| | |||| |||||| ||| |||| || ||||||| | || | || |||||||
Db   20 CAGTGCCTGAACCAAGAGGCATTGGAGCCGCTATGACGGCGTTGTCCCAAGGCAAGTACACT 79

Qy   289 ATTGGCTTGGGCCAGGCCAAGATAGGGCTTTCTCGACAGATAGAGAAGAATATTAACTCTCTT 348
          | |||| |||| ||| |||| || ||| ||| | |||| ||| ||| | || || ||||
Db   80 ATCGGCTTGGGCCGACAACAACATGAGCTTTGTGAACGACCACGCGAGGACAATCTACTCGATG 139

Qy   349 TGCATGACTGTGTGGTTCAGAATCTTATGGGAGGAGAAGAAATAAACCTTTCCTATATGATTCCATTGGG 408
          |||| | | | |||||||| | || ||||| |||| || | ||| ||| || ||||| |||
Db   140 TGTTTGACCCGGTGCTCGGAACTGATGAAGAACTAGGACTATCACCCGGAGAAGACATCGGC 199

Qy   409 CGGCTCGGAAGTTGGAACAGAGAGACAAATCATCGACAAAATCAAAGTCTGTGAAGACTAATTG 468
          ||||| |||| || | ||||| | || || || ||| |||| | |||| ||| || |
Db   200 CGCCTCGAGTGGGTACGGAGACGTTGCCTTGACAAGTCGAAGTCCGTGAAGTCTATTTG 259

Qy   469 ATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATAGAAGGAATCGACACAACTAATGCA 528
          || | || ||| ||||| ||||| |||| |||| | |||| || || | | ||| |||
Db   260 AATGCAGTTGTTCTGCGGCGA------GAACACCGACTTGGAGGGTGTCGATACCTGAACGGCC 313

Qy   529 TGCTATGGAGGACGCAGCTGCTGTCTTCAATCGCTGTTAACTGGAATTGAGTCCAGCTCTTGG 588
          |||| ||| || | ||| || ||| | || ||| ||| ||||| ||| |||| | |||
Db   314 TGCTATGGGCGGTACTAACGCGTTGTTTAACTCCTTGAACTGGATTGAGTTCGGTGG 373

Qy   589 GATGGGCTT 597
          ||| || |
Db   374 GACGGTCGT 382

; TITLE OF INVENTION: No. 6672186e1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 744
; LENGTH: 307
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(307)
; OTHER INFORMATION: n = A,T,C or G

US-09-833-381-744

Best Local Similarity 56.9%;   Pred. No. 9.9e-18;   Indels 131;   Gaps 0;
Query Match   5.5%;   Score 110;   DB 4;   Length 307;
Matches 173;   Conservative 0;   Mismatches 131;   Indels 0;   Gaps 0;

Qy   775 TTTACTTGAATGATTTGGCTTCATGATCTTTCACTCACCCATATTGTAAAACTGGTTCAG 834
          ||| | || ||||||| | ||| || || | | || | || || || ||| | |
Db   2 TNCACCCTNGACGATTACAGTACATGANCTTTCATCACNCCTTTTGCAAGANNGTCCAG 61

Qy   835 AAATCTCTAGCTCGATGTTGCTGAATGACTTCCTTAATGACACAGAATAGAGATAAAAT 894
          || | | || | | ||| | || || | | || | || |||| || || |
Db   62 ANGNCTCTNGCTCNGCTGATNTTCAATGACTTNCTGTCAGCNAGCAGTGACACACAAACC 121

Qy   895 AGTATCTATAGTGGCCCTGGAAGCCTTTGGGGATGGTTAAATTAGAAGACACCTACTTTGAT 954
          | |||| ||| ||| ||| ||||| | | ||| ||||| ||||| ||| | || |||
Db   122 AGCTTATATAAGGGCTNGANGGCTTTCGGGGGGCTNAAAGCTGGANGACACCTNCACCAAC 181

Qy   955 AGAGATGTGGAGAGGCATTTATGAAGGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAG 1014
          |||| ||||| ||| ||| | |||| ||| |||| || |||| ||| |||||| || |||
Db   182 AAGGACTCGGNTAAAGCACTTCTANAGGCCTNTNAGGACACATGTTCGACAAGANANACCAAG 241

Qy   1015 GCCATCTTTACTTGTTATCAAATCAAAATTGGAAAATAATGTACACATCTTCAGTTATTGGTTCC 1074
          || | | || | | | || || || || || ||| | |||| | ||| ||| | ||| ||
Db   242 GCTTCCCTTTACCTCTCCACTCACAAATNGAACATGTTGNCANCTCNACCCCTGTTACGGGGTNC 301

Qy   1075 CTTG 1078
          ||||
Db   302 CTGG 305

GENERAL INFORMATION:
;   APPLICANT: HOSHINO, Tatsuo
;   APPLICANT: OJIMA, Kazuyuki
;   APPLICANT: SETOGUCHI, Yutaka
;   TITLE OF INVENTION: ISOPRENOID PRODUCTION
;   FILE REFERENCE: ISOPRENOID PRODUCTION
;   CURRENT APPLICATION NUMBER: US/09/925,388
;   CURRENT FILING DATE: 2001-08-09
;   PRIOR APPLICATION NUMBER: 09/306,595
;   PRIOR FILING DATE: 1999-05-06
;   NUMBER OF SEQ ID NOS: 43
;   SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1
;   LENGTH: 4775
;   TYPE: DNA
;   ORGANISM: Phaffia rhodozyma
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: (1239)..(1240)
;   OTHER INFORMATION: EXPERIMENTAL
;   NAME/KEY: exon
;   LOCATION: (1305)..(1361)
;   NAME/KEY: intron
;   LOCATION: (1362)..(1504)
;   NAME/KEY: exon
;   LOCATION: (1505)..(1522)
;   NAME/KEY: intron
;   LOCATION: (1523)..(1699)
;   NAME/KEY: exon
;   LOCATION: (1700)..(1826)
;   NAME/KEY: intron
;   LOCATION: (1827)..(1920)
;   NAME/KEY: exon
;   LOCATION: (1921)..(2277)
;   NAME/KEY: intron
;   LOCATION: (2278)..(2351)
;   NAME/KEY: exon
;   LOCATION: (2352)..(2409)
;   NAME/KEY: intron
;   LOCATION: (2410)..(2497)
;   NAME/KEY: exon
;   LOCATION: (2498)..(2504)
;   NAME/KEY: intron
;   LOCATION: (2505)..(2586)
;   NAME/KEY: exon
;   LOCATION: (2587)..(2768)
;   NAME/KEY: intron
;   LOCATION: (2769)..(2851)
;   NAME/KEY: exon
;   LOCATION: (2852)..(2891)
;   NAME/KEY: intron
;   LOCATION: (2892)..(2985)
;   NAME/KEY: exon
;   LOCATION: (2986)..(3240)
;   NAME/KEY: intron
;   LOCATION: (3241)..(3325)
;   NAME/KEY: exon
;   LOCATION: (3326)..(3493)
;   NAME/KEY: intron
;   LOCATION: (3494)..(3601)
;   NAME/KEY: exon
;   LOCATION: (3602)..(3768)
;   NAME/KEY: polyA_site
;   LOCATION: (4043)..(4044)
US-09-306-595C-1

Query Match   5.2%; Score 103.6; DB 3; Length 4775;
Best Local Similarity   63.0%; Pred. No. 1.1e-15;
Matches 160; Conservative   0; Mismatches 94; Indels   0; Gaps   0;

RESULT 14
US-09-925-388-1
; Sequence 1, Application US/09925388
; Patent No. 6586202

Qy   340 AACTCTCTTTGCATGACTGTGGTTCAGAAATCTTATGGAGAGAAATAACCTTTCCTATGAT 399
Db   1902 AATTCCATTTGATTGTTAGCTGTTTCCGGTCTTCTATCAAAGTACAACGTTGATCCCAAG 1961

Qy   400 TGCATTGGGCGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAAATCAAAGTCTGTGAAG 459
Db   1962 TCAATCGGTCGGAATTGATGTCGGAACTGAGTCCATCATCATTGACAAGTCCAAAATCTGTGAGTCG 2021

Qy   460 ACTAAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATAGAGAGGAATCGACACA 519
Db   2022 ACAGTCCTTATGGACTTGTTCGAGTCCCACGGCAACACAGATATTGAGGGTATCGACTCC 2081

Qy   520 ACTAAATGCATGCTATGGAGGCACAGCTGCTGTCTGTTCAATGCTGTTAACTGGGATTGAGTCC 579
Db   2082 AAGAATGCCTGCTACGGTTCTACCGCGCGGCCCTGTTCAATGCCCGTCAACTGGGATCGAGTCA 2141

Qy   580 AGCTCTTGGGATGG 593
Db   2142 TCCTCTTGGGACGG 2155

Query Match   5.2%; Score 103.6; DB 4; Length 4775;
Best Local Similarity   63.0%; Pred. No. 1.1e-15;
Matches 160; Conservative   0; Mismatches 94; Indels   0; Gaps   0;
US-09-925-388-1

Qy   340 AACTCTCTTTGCATGACTGTGGTTCAGAAATCTTATGGAGAGAAATAACCTTTCCTATGAT 399
Db   1902 AATTCCATTTGATTGTTAGCTGTTTCCGGTCTTCTATCAAAGTACAACGTTGATCCCAAG 1961

```
Qy   400 TGCAATTGGCGCGGCCTGGAAGTTGGAACAGAGACAATCATCGACAAAATCAAAGTCTGTGAAG 459
         ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1962 TCAATCGGTCGGAATTGATCGTCGGAACTGAGTCCATCATTGACAACAGTCCAAAATCTGTCAAG 2021

Qy   460 ACTAATTTGATGCAGCTCGTTTGAAGAGTCTCGGGAATACAGATATAGAAGGAATCGACACA 519
         ||||||| |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db  2022 ACAGTCCTTATGGACTTGTTCGAGTCCCACGCACACAGATATTGACGGTATCGACTCC 2081

Qy   520 ACTAATGCAATGCTATGAGGCACAGCTCGTGTCTTCAAATGCTGTTTAACTGGATTGAGTCC 579
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2082 AAGAATGCCCGCTACGGTTCTACCGACGCCCCCGTTCAAATCGCCGTCAACTGGATCCGAGCTCA 2141

Qy   580 AGCTCTTGGGATGG 593
         |||||||||||||
Db  2142 TCCTCCTTGGGACGG 2155
```

RESULT 15
US-09-313-294A-4313
; Sequence 4313, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 4313
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6476212 7003482141H1
US-09-313-294A-4313

Query Match   4.3%; Score 86.6; DB 4; Length 305;
Best Local Similarity   66.1%; Pred. No. 5.3e-12;
Matches 125; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
Qy   414 GGAGTTGGAACAGAGACAATCATCGACAAAATCAAAGTCTGTGAAGACTAATTTGATGCA 473
         |||||||||| |||||||||| |||||| ||||||| | ||||||||||||||||| |||||
Db     1 GGAGGTTGGAAGTCGAAAACAGTTATAGACAAAAGTAAAATCCATCAAAAACTTGGCTGATGCA 60

Qy   474 GCTCGTTTGAAGAGTCTCGGGAATACACATATAGAAGGAATCGACACAACTAATGCACATGCTA 533
         |||||| |||||||||||||| ||| || || || |||| ||| |||| |||||| |||||||||
Db    61 GATCTTTGAGGAAAGTGTAACAGTGACATTCGAAGGTGTTGACTTCCAGCAATCCATCTTA 120

Qy   534 TGGAGGCACAGCTCGTGTCTTCAATGCTGTTTAACTGGATTGAGTCCAGCTCTTGGGATGG 593
         ||||| |||| |||||   ||| |||||| |||| ||||||||| || || ||||||||
Db   121 TGGCCGGGACAGCAGCCCTACTGAAATTCGTGTGAATTGGGTCGAAAGTAATTCCTGGGATGG 180

Qy   594 GCTTCGTGG 602
         |||||||||
Db   181 CCGTTATGG 189
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: June 24, 2004, 09:40:27 ; Search time 840 seconds
(without alignments)
10918.263 Million cell updates/sec

Title: 2002
Perfect score: 2002
Sequence: 1 cgcctcccagcgactcctcgg.........aaaaaaaaaaaaaaaaaaaaaa 2002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

```
 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 3: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
 5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 6: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 7: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2002	100.0	2002	14	US-10-193-295-1	Sequence 1, Appli
2	2002	100.0	2002	16	US-10-622-516-1	Sequence 1, Appli
3	1693	84.6	3722	10	US-09-919-039-98	Sequence 98, Appl
4	1370	68.4	1650	13	US-10-236-417-181	Sequence 181, App
5	1370	68.4	1650	13	US-10-236-417-183	Sequence 183, App
6	1370	68.4	1650	13	US-10-236-417-185	Sequence 185, App
7	1370	68.4	1650	13	US-10-236-417-189	Sequence 189, App
8	1370	68.4	1650	13	US-10-236-417-193	Sequence 193, App
9	1370	68.4	1650	13	US-10-236-417-197	Sequence 197, App
10	1370	68.4	1650	13	US-10-236-417-201	Sequence 201, App
11	1370	68.4	1650	13	US-10-236-417-205	Sequence 205, App
12	1370	68.4	1650	13	US-10-307-817-359	Sequence 359, App
13	1370	68.4	1650	13	US-10-307-817-371	Sequence 371, App
14	1370	68.4	1650	13	US-10-307-817-445	Sequence 445, App
15	1370	68.4	1650	15	US-10-354-358-55	Sequence 55, Appl
16	1322.4	66.1	3008	15	US-10-106-698-1926	Sequence 1926, Ap
17	1302	65.0	1564	13	US-10-236-417-203	Sequence 203, App
18	1302	65.0	1564	13	US-10-236-417-207	Sequence 207, App
19	1302	65.0	1564	13	US-10-307-817-369	Sequence 369, App
20	1302	65.0	1564	13	US-10-307-817-373	Sequence 373, App
21	1299.4	64.9	1601	13	US-10-236-417-191	Sequence 191, App
22	1299.4	64.9	1601	13	US-10-307-817-363	Sequence 363, App
23	1299	64.9	1593	13	US-10-236-417-187	Sequence 187, App
24	1299	64.9	1593	13	US-10-307-817-361	Sequence 361, App
25	1299	64.9	1612	13	US-10-236-417-199	Sequence 199, App
26	1299	64.9	1612	13	US-10-307-817-357	Sequence 357, App
27	1299	64.9	1612	13	US-10-307-817-367	Sequence 367, App
28	1295.4	64.7	1612	13	US-10-236-417-195	Sequence 195, App
29	1295.4	64.7	1608	13	US-10-307-817-365	Sequence 365, App
30	1295.4	64.7	1608	13	US-10-236-417-211	Sequence 211, App
31	1295.4	64.7	1608	13	US-10-307-817-375	Sequence 375, App
32	1295	64.7	1608	13	US-10-307-817-365	Sequence 365, App
33	1180.6	59.0	1608	13	US-10-388-934-37	Sequence 37, Appl
34	1180.6	59.0	3275	16	US-10-388-934-37	Sequence 37, Appl
35	1103.2	55.1	3275	16	US-10-205-331-15	Sequence 15, App
36	1103.2	55.1	170834	9	US-09-835-232-7	Sequence 7, Appli
37	476	23.8	170834	15	US-09-308-485-7	Sequence 7, Appli
38	476	23.8	28001	14	US-10-193-295-3	Sequence 3, Appli
39	460.4	23.0	28001	16	US-10-622-516-3	Sequence 3, Appli
40	446	22.3	565	9	US-09-833-790-10	Sequence 10, Appl
41	446	22.3	849	13	US-10-027-632-164650	Sequence 164650,
42	446	22.3	849	13	US-10-027-632-164651	Sequence 164651,
43	446	22.3	849	13	US-10-027-632-164652	Sequence 164652,
44	446	22.3	849	16	US-10-027-632-164651	Sequence 164651,
45	446	22.3	849	16	US-10-027-632-164652	Sequence 164652,

ALIGNMENTS

RESULT 1
US-10-193-295-1

Sequence 1, Application US/10193295

Publication No. US20020173018A1

GENERAL INFORMATION:
APPLICANT: GONG, Fangcheng et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL001119SDIV
CURRENT APPLICATION NUMBER: US/10/193,295
PRIOR APPLICATION NUMBER: 08/819,993
PRIOR APPLICATION DATE: 2001-03-29
CURRENT FILING DATE: 2002-07-12
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2002
TYPE: DNA
ORGANISM: Human
US-10-193-295-1

Query Match 100.0%; Score 2002; DB 14; Length 2002;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CGCCTCCCAGCGACTCCTCGGCCAGTGCCGGAGTCGGGTGGGGTTGGGCGGCTATAAAGCTGGT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCCTCCCAGCGACTCCTCGGCCAGTGCCGGAGTCGGGTGGGGTTGGGCGGCTATAAAGCTGGT 60

Qy     61 AGCGAAGGGGGAGGGCCCCGCGGCCGCGCGGCACTCCTTCGTTGGTGCTCACTCCCTTTCTCTGCCG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGCGAAGGGGGAGGGCCCCGCGGCCGCGCGGCACTCCTTCGTTGGTGCTCACTCCCTTTCTCTGCCG 120

Qy    121 CTCGGGTCACGCGTTGCTTCTTTCACCATGCCTGGGATCACTTCCTTTGAATGCAGAAGCTTGC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCGGGTCACGCGTTGCTTCTTTCACCATGCCTGGGATCACTTCCTTTGAATGCAGAAGCTTGC 180
```

Qy   181 TGGGCCAAAAGAATGTGGGAATTGTTGCCCTTGGAGATCTATTTTCCTTCTCAATATGTTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGGCCAAAAGAATGTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGAT 240

Qy   241 CAAGCAGGAGTTGGAAAAATATGATGGTTGTAGATGCTGGAAAGTATACCATTGGCTTGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAACCAGAGTTGGAAAAATATGATGGTTGTAGATGCTGGAAAGTATACCATTGGCTTGGGC 300

Qy   301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAGATATTAACTCTCTTTGCATGACTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CAGGCCAAGATGGGCTTCTGCACAGATAGAGAGATATTAACTCTCTTTGCATGACTGTG 360

Qy   361 GTTCAGAATCTTATGGAGGAAAATAACCTTTCCTATGATGATTGCATTGGGCGGCTGGAAGTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTTCAGAATCTTATGGAGGAAAATAACCTTTCCTATGATGATTGCATTGGGCGGCTGGAAGTT 420

Qy   421 GGAACAGAGACAATCATCGACAGAAATCAAAGTCTGTGAAGACTAAATTTGTTGATGGCAGCTGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GGAACAGAGACAATCATCGACAGAAATCAAAGTCTGTGAAGACTAAATTTGTTGATGGCAGCTGTTT 480

Qy   481 GAAGAGTCTGGGAATACAGATAATAGAGGAATCGACACAACTAATGCATGCTTATGGAGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GAAGAGTCTGGGAATACAGATAATAGAGGAATCGACACAACTAATGCATGCTTATGGAGGC 540

Qy   541 ACAGCTGCTGTCTTCTTCAAATGCTGTTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACAGCTGCTGTCTTCAAATGCTGTTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600

Qy   601 GGGACACATATGCAACATGCCTTATGTATGATATGCTATCTGAAATATCCT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGGACACATATGCAACATGCCTTATGTATGATATGCTATCTGAAATATCCT 660

Qy   661 ATAGTAGATTAGGGAGAAACTCTCCATACACAGTGCTACCTTCACAGTGCATTGCTACTCTCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATAGTAGATGGGAAAACTCTCCATACACAGTGCTACCTTCACAGTGCATTGCTACTCTCT 720

Qy   721 GTCTACTGCAAAAAGATCCATCCCCAGTGGGCAGAAAAGAGGGAAATGATAAAAGATTTTACC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 GTCTACTGCAAAAAGATCCATCCCCAGTGGGCAGAAAAGAGGGAAATGATAAAAGATTTTACC 780

Qy   781 TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATAGTGTAAACTGGTTCAGAAATCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TTGAATGATTTTGGCTTCATGATCTTTCACTCACCATAGTGTAAACTGGTTCAGAAATCT 840

Qy   841 CTAGCTCGGATGTTGCTCGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CTAGCTCGGATGTTGCTCGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATC 900

Qy   901 TATAGTAGGGCCTGGAAGCCTTTGGGGATGTTAAATATTAGAAGACACCTACTTTGATAGAGAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TATAGTGGGCCTGGGAAGCCCTTTGGGGAGTGTTAAATATTAGAAGACACCTACTTTGATAGAGAT 960

Qy   961 GTGGGAGAAGGCATTTATGAAGGCTAGCTCTGAACTTCTTCAGTCAGAAAACAAAGGCATCT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GTGGGAGAAGGCATTTATGAAGGCTAGCTCTGAACTTCTTCAGTCAGAAAACAAAGGCATCT 1020

Qy   1021 TTACTTGTATCAAAATCAAAATGGAAAATAATGTACACATCTTCTTCAGTGTAATGGTTCCCTTGCA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACTTGTATCAAAATCAAAATGGAAAATAATGTACACATCTTCAGTATATGGTTCCCTTGCA 1080

Qy   1081 TCTGTTTCTAGCCACAGTACTCACCTCACCTCCAGCAATTAGCAGGGAGGAAGGAGGAGTGTTTTTCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTGTTTCTAGCCACAGTACTCACCTCACCTCCAGCAATTAGCAGGGAGGAAGGAGGAGTGTTTTTCT 1140

Qy   1141 TATGGTTCTGGTTTTGGCTGCCACTCTCTGTACTCTCTTCTTTAAAGTCACACAAGAATGCTACACCG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TATGGTTCTGGTTTTGGCTGCCACTCTCTGTACTCTCTTCTTTAAAGTCACACAAGAATGCTACACCG 1200

Qy   1201 GGGTCTGCCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCA 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGTCTGCCTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCA 1260

Qy   1261 AGAACTGGTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGAGGACACCCAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAACTGGTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGAGGACACCCAT 1320

Qy   1321 CATTTGGGTCAAACTATATATTCCCCAGGGGTTCAATAGAGTTCACTCTTTTGAAGGAACGTTGGTAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CATTTGGGTCAAACTATATATTCCCCAGGGGTTCAATAGAGTTCACTCTTTTGAAGGAACGTTGGTAC 1380

Qy   1381 TTAGTTAGGGTGGATGAAAAAGCACAGCAGAAGAACACTTACGCTCGGGCGTCCCACTCCCAAATGAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTAGTTAGGGTGGATGAAAAAGCACAGCAGAAGAACACTTACGCTCGGGCGTCCCACTCCCAAATGAT 1440

Qy   1441 GACACTTTGGATGAAGAGAGTAGGACTTGTGCCATTCAAAAATGACAAACTGACAATATTCCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GACACTTTGGATGAAGAGAGTAGGACTTGTGCCATTCAAAAATGACAAACTGACAATATTCCA 1500

Qy   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTC 1560

Qy   1561 ATTAGTAAATGGGGAACAATTAAGATACTCTGTTGAGGTGCAAGAACTTCAGGGGTGGGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATTAGTAAATGGGGAACAATTAAGATACTCTGTTGAGGTGCAAGAACTTCAGGGGTGGGGC 1620

Qy   1621 ATGGGGGTGGGGGGTATGGGAACAGTTGGACGGGAATCGGGGATAATCTGGGGATAAATTTTAAAGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGGGGTGGGGGGTATGGGAACAGTTGGACGGGAATCGGGGATAATCTGGGGATAAATTTTAAAGGA 1680

Qy   1681 TTACATGTTATGTAAATTTTTATGTGACTGACATGGAGCTGGAGCTGGATGACATATGCGTACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTACATGTTATGTAAATTTTTATGTGACTGACATGGAGCTGGATGACATATGCGTACTT 1740

Qy   1741 GGGAAAGTCTCTTTTGCTCGTATTTTGCTGAACATGCTTCCGTTGTTGTGTGGCCAATGCCAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGGAAGAAGTCTCTTTTGCTCGTATTTTGCTGAACATGCTTCCGTTGTTGTGTGGCCAATGCCAA 1800

Qy   1801 ATGTACTCGAATGATGTTAAGGGCTCTTGTAAAACTTCATACCTCTTCGTTGGCCATTTGTATG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATGTACTCGAATGATGTTAAGGGCTCTTGTAAAACTTCATACCTCTTCGTTGGCCATTTGTATG 1860

Qy   1861 CATGATGTTTGGTTTTTAAACATGGTATAATCGATATTGTGTACTTCTGTCAGAAGAAAGCA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CATGATGTTTGGTTTTTAAACATGGTATAATCGATATTGTGTACTTCTGTCAGAAGAAAGCA 1920

Qy   1921 GAGGTACTAATCTCCAATTAAAAAATTTTTTTAAACATGTAAAAAAAAAAAAAAAAAAAAAA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGGTACTAATCTCCAATTAAAAAATTTTTTTAAACATGTAAAAAAAAAAAAAAAAAAAAAA 1980

Qy   1981 AAAAAAAAAAAAAAAAAAAAAAAA 2002
          ||||||||||||||||||||||||
Db   1981 AAAAAAAAAAAAAAAAAAAAAAAA 2002

```
; TYPE: DNA
; ORGANISM: Human
US-10-622-516-1

Query Match    100.0%;  Score 2002;  DB 16;  Length 2002;
Best Local Similarity  100.0%;  Pred. No. 0;  Mismatches 0;
Matches 2002;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      1 CGGCTCCCAGCGACTCTCGGACTCGCCAGTCGCCGGAGTCGGGTCGGGGGGCTATAAAGCTCGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGCTCCCAGCGACTCTCGGACTCGCCAGTCGCCGGAGTCGGGTCGGGGGGCTATAAAGCTCGGT 60

Qy     61 AGGGAAGGGGAGGGCGGCCGGCCCCGGGGACTGTCCTTCCTTCCTCCACTCCCTTCCTCTGCCG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGGAAGGGGAGGGCGGCCGGCCCCGGGGACTGTCCTTCCTTCCTCCACTCCCTTCCTCTGCCG 120

Qy    121 CTCGGTCACGCTTGCTCTCTTTCACCATGCCTGGATCACTTCCTTCCTTGAATGCAGAAGCTTGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCGGTCACGCTTGCTCTCTTTCACCATGCCTGGATCACTTCCTTCCTTGAATGCAGAAGCTTGC 180

Qy    181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCTCAATATGGTTGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGCCAAAAGATGTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCTCAATATGGTTGAT 240

Qy    241 CAAGCAGAGTTGGAAAAATATGAGTGGTGGAGAGGCTGGGAAAGTATACATTGGCTTGGGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAAGCAGAGTTGGAAAAATATGAGTGGTGGAGAGGCTGGGAAAGTATACATTGGCTTGGGC 300

Qy    301 CAGGCCAAGAGATGGGGCTTCTGCACAGATAGGAGAGAAGAAGATATTAACTCTCTTCTTGCATGACTGTG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGCCAAGAGATGGGGCTTCTGCACAGATAGGAGAGAAGAAGATATTAACTCTCTTCTTGCATGACTGTG 360

Qy    361 GTTCAGAATCTTATGGAGGAGAAATAACCTTTCCTATGATTGCATTGGGGCGGCTTGGAAGTT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTTCAGAATCTTATGGAGGAGAAATAACCTTTCCTATGATTGCATTGGGGCGGCTTGGAAGTT 420

Qy    421 GGAACAGAGACAATCATCGACAAATCAAAAGTCTGTGAAGACTAAATCTGAAGTGCAGGCTGTTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGAACAGAGACAATCATCGACAAATCAAAAGTCTGTGAAGACTAAATCTGAAGTGCAGGCTGTTT 480

Qy    481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGGACACAACAAACTAATGCTATGGGAGGC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAAGAGTCTGGGAATACAGATATAGAAGGAATCGGACACAACAAACTAATGCTATGGGAGGC 540

Qy    541 ACAGCTGCTGTCTTCTTCAATCGTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ACAGCTGCTGTCTTCTTCAATCGTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGT 600

Qy    601 GGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATCTGAATATCTT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGATATCTGAATATCTT 660

Qy    661 ATAGTAGATGGAAAACTCTCCATACAGTGCCACAGTTCAGTTGCATTAGACCGCTGCTACTCT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATAGTAGATGGAAAACTCTCCATACAGTGCCACAGTTCAGTTGCATTAGACCGCTGCTACTCT 720

Qy    721 GTCTACTGCGAAAAAGATCCATGCCCAGTGGCGCAGAAAGAGGAAATGATAAAGATTTTACC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTCTACTGCGAAAAAGATCCATGCCCAGTGGCGCAGAAAGAGGAAATGATAAAGATTTTACC 780

Qy    781 TTGAATGATTTGGCTTCACTCACCATCACCATATTGTTAAACTGGTTCAGAAAATCT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TTGAATGATTTGGCTTCACTCACCATCACCATATTGTTAAACTGGTTCAGAAAATCT 840

Qy    841 CTAGCTCGGATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAACCAGAAATTAGTAGTATC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTAGCTCGGATGTTGCTGAAATGACTTCCTTAATGACCAGAATAGAACCAGAAATTAGTAGTATC 900

Qy    901 TATAGTGGCTGGAAGCCTTTGGGGGATGTTAAAATTAGAAGACACACCTACTTTGATAGAGAT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TATAGTGGCTGGAAGCCTTTGGGGGATGTTAAAATTAGAAGACACACCTACTTTGATAGAGAT 960

Qy    961 GTGGAGAGGCATTTATGAAGGCTAGCTCTGACTCTTCAGTCAGATATATGTTCCCTTGCA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTGGAGAGGCATTTATGAAGGCTAGCTCTGACTCTTCAGTCAGATATATGTTCCCTTGCA 1020

Qy   1021 TTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTGTATATGGTTTCCTTTCT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TTACTTGTATCAAATCAAAATGGAAATATGTACACATCTTCAGTGTATATGGTTTCCTTTCT 1080

Qy   1081 TCTGTTCTAGCACAGTACTCACCTCAGTCCAGCAATTAGCCAGGGGAAGGAAGGAAGGAGATGGAGTGTGTTCA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCTGTTCTAGCACAGTACTCACCTCAGTCCAGCAATTAGCCAGGGGAAGGAAGGAAGGAGATGGAGTGTGTTCA 1140

Qy   1141 TATGGTTCGGTTGGCTGCCGACTCCTGTGACTCGTACTCTCTTAAAGTCACAACAAGATGCTACACCG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TATGGTTCGGTTGGCTGCCGACTCCTGTGACTCGTACTCTCTTAAAGTCACAACAAGATGCTACACCG 1200

Qy   1201 GGGTCTGGCTCTTGATGAAAATAACACGAAGTTTATGTTTATGTGATCTTAAATCAAGGCTTGGGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGGTCTGGCTCTTGATGAAAATAACACGAAGTTTATGTTTATGTGATCTTAAATCAAGGCTTGGGC 1260

Qy   1261 AGAACTGGTGTCGGCACCAGAGTGTCTTCGCTGGAAAAACATGAAGCTTCAGAGGACACCCAT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAACTGGTGTCGGCACCAGAGTGTCTTCGCTGGAAAAACATGAAGCTTCAGAGGACACCCAT 1320

Qy   1321 CATTTGGTCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAAACGTGGTAC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CATTTGGTCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAAACGTGGTAC 1380

Qy   1381 TTAGTTAGGGTGGATGAAGAACATTAGAATACTCTGTGAGGTGCAAGAGAACTTCAGGGTGGGC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTAGTTAGGGTGGATGAAGAACATTAGAATACTCTGTGAGGTGCAAGAGAACTTCAGGGTGGGC 1440

Qy   1441 GACACTTTGGATGAAGGGAGGACTTGTTGCATTCAAACATAGCAACATAGCATATTCCA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GACACTTTGGATGAAGGGAGGACTTGTTGCATTCAAACATAGCAACATAGCATATTCCA 1500

Qy   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTC 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTC 1560

Qy   1561 ATTAGTAATGGGGACATTTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGGGC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATTAGTAATGGGGACATTTAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGGGC 1620

Qy   1621 ATGGGGGTGGGGGTATGGGGAACAGTTGGGAATGGGGATATCTGGGGATAATTTTAAAGGA 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGGGGTGGGGGTATGGGGAACAGTTGGGAATGGGGATATCTGGGGATAATTTTAAAGGA 1680

Qy   1681 TTACATGTTATGTAAAATTTTTATGTGACTGACATGGACGATGACTAGCTACTT 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TTACATGTTATGTAAAATTTTTATGTGACTGACATGGACGATGACTAGCTACTT 1740

Qy   1741 GGGAAAGTCTCTTTGCTCTATTTGCTGACACATGCTTCCTGTTGTGGTCTGGCCAATGCCAA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GGGAAAGTCTCTTTGCTCTATTTGCTGACACATGCTTCCTGTTGTGGTCTGGCCAATGCCAA 1800

Qy   1801 ATGTACTCGAATGATGTTAAGGGCTCTGTAAAACTTCATCTCTTCTTTGGGCCATTTGTATG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATGTACTCGAATGATGTTAAGGGCTCTGTAAAACTTCATCTCTTCTTTGGGCCATTTGTATG 1860

Qy   1861 CATGATGTTTGGTTTTTAAACATGGTATATATGTAATGAATTGTACTTCGTCAGAAGAAAGCA 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CATGATGTTTGGTTTTTAAACATGGTATATATGTAATGAATTGTACTTCGTCAGAAGAAAGCA 1920

Qy   1921 GAGGTACTAATCTCCAATTAAAAAATTTTTAACATGTAAAAAAAAAAAAAAAAAAA 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 GAGGTACTAATCTCCAATTAAAAAATTTTTAACATGTAAAAAAAAAAAAAAAAAAA 1980

Qy   1981 AAAAAAAAAAAAAAAAAAAAAAA 2002
          ||||||||||||||||||||||||
Db   1981 AAAAAAAAAAAAAAAAAAAAAAA 2002
```

RESULT 3

```
US-09-919-039-98
; Sequence 98, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
;   APPLICANT: Kaser, Matthew R.
;   TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
;   FILE REFERENCE: PA-0035 US
;   CURRENT APPLICATION NUMBER: US/09/919,039
;   CURRENT FILING DATE: 2002-09-09
;   PRIOR APPLICATION NUMBER: 60/222,113
;   PRIOR FILING DATE: 2000-07-28
;   NUMBER OF SEQ ID NOS: 401
;   SOFTWARE: PERL Program
; SEQ ID NO 98
;   LENGTH: 3722
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030108871A1 347055.4
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 104, 2838
;   OTHER INFORMATION: a, t, c, g, or other
US-09-919-039-98

Query Match             84.6%;  Score 1693;  DB 10;  Length 3722;
Best Local Similarity   92.7%;  Pred. No. 0;  Mismatches 20;  Indels 126;  Gaps 1;
Matches 1841;  Conservative    0;  Mismatches  20;  Indels  126;  Gaps  1;

Qy      134 GCTCTTCACCATGCCTGGATCACTTCCTTTCCTTTGAATCGCAGAAGCTTGCTTGCCCAAAAGATG 193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      238 GCTCTTCACCATGCCTGGATCACTTCCTTTCCTTTGAATCGCAGAAGCTTGCTTGCCCAAAAGATG 297

Qy      194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      298 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 357

Qy      254 AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      358 AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 417

Qy      314 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGCATGACTGTGGTTCAGAATCTTA 373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      418 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 477

Qy      374 TGGGAGAGAAATAACCTTTCCTATCATTGCATTGGGCGCGCCTGGAAGTGGAACACAGAGACAA 433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      478 TGGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGCGCCTGGAAGTGGAACACAGAGACAA 537

Qy      434 TCATCCGACAAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGCA 493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      538 TCATCGACAAAATCAAAGTCTGTGAAGACTAATTGATGCAGCTGTTTGAAGAGTCTGGGA 597

Qy      494 ATACAGATATAGAAGGAATGACACAATCATATCATTGCTATGGAGGCACAGCTGCTGCTCT 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      598 ATACAGATATAGAAGGAATGACACAACTAATGCATGGAGGCACAGCTGCTGCTCT 657

Qy      554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----------------- 591
            |||||||||||||||||||||||||||||||||||||||
Db      658 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGAGCGTATGCCCTGTGAGTTG 717

Qy      592 ---------- 591
            
Db      718 CAGGAGATAATTGCTGTATATGCCACAGAGAAATGCTAGACCTACAGGTGGAGTTGGAGCAG 777

Qy      592 --------------------GGGCTTCGTGGGGACAC 607
                                ||||||||||||||||
Db      778 TAGCTCTGCTAATTGGGCAAATGCTCCTTTTAATTTTTGAAGGAGGGCTTCGTGGGGACAC 837

Qy      608 ATATGCAACATGCCTATGAATTTTTACAAGCCTTGATAATGCTATCTCGAATATCCTATAGTAG 667
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      838 ATATGCAACATGCCTATGAATTTTTACAAGCCTTGATAATTTTTACAAGCCTATAGTAG 897

Qy      668 ATGGAAAACTCTCCAATACAGTGCTACCTCAGTGCATTCAGTGCATTAGAGA 727
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      898 ATGGAAAACTCTCCAATACAGTGCTACCTCAGTGCATTCAGTGCATTAGACCGCTGCTCTCCTGTCTACT 957

Qy      728 GCAAAAAGATCCATGCCCAGTGGCAGAAGAGGGAAATGATTAGATTTAACCTTGAATG 787
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      958 GCAAAAAGATCCATGCCCAGTGCAGAAGAGGGAAATGATTAGATTTAACCTTGAATG 1017

Qy      788 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTAAATCGGTTCAGAAATCTCAGTG 847
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1018 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTAAATCGGTTCAGAAATCTCAGTG 1077

Qy      848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAAATAGTATCTATAGTG 907
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1078 GGATGTTGCTGAATGACTTCCTTAATGACCAGAAATAGAGATAAAAATAGTATCTATAGTG 1137

Qy      908 GCCTGGAAGCCTTTGGGGATGTTAAAATTAGAAGAACACCTACTTTGATAGAGATGTGGAGA 967
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1138 GCCTGGAAGCCTTTGGGGATGTTAAAATTAGAAGAACACCTACTTTGATAGAGATGTGGAGA 1197

Qy      968 AGGCATTTATGAAGGCTAGCTTCTTGAACTCTTCAGTGAAACTTCAGTGAAAAAGGCATCTTTACTTG 1027
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1198 AGGCATTTATGAAGGCTAGCCTTTGAACTCTTCAGTGAAACTTCAGTGCAGAAAAACAAAGGCATCTTTACTTG 1257

Qy     1028 TATCAAAATCAAAAAATGGAAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC 1087
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1258 TATCAAAATCAAAAAATGGAAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC 1317

Qy     1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGAGCAGAGAAATTGGGAGTGTTTTTCTTATGGTT 1147
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1318 TAGCACAGTACTCACCTCAGCAATTAGCAGGAGCAGAGAAATTGGGAGTGTTTTTCTTATGGTT 1377

Qy     1148 CTGGTTTGGCTGCCACTCTTGTACTCTCCTTTAAAGTCACACAAGATGCTCACACCGGGGTCTG 1207
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1378 CTGGTTTGGCTGCCACTCTTGTACTCTCCTTTAAAGTCACACAAGATGCTCACACCGGGGTCTG 1437

Qy     1208 CTCTTGATAAAATAACACGAAGTTTATGTGATCTTTAAATCAAGGCTTGATTCAAGAACTG 1267
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1438 CTCTTGATAAAATAACACGAAGTTTATGTGATCTTTAAATCAAGGCTTGATTCAAGAACTG 1497

Qy     1268 GTGTTGGCACCAGAGTGTCTTCGCTGTGAAAAACATGAAGCTCAGAGGAGCACACCCATTTGG 1327
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1498 GTGTTGGCACCAGAGTCTTCGCTGTGAAAAACATGAAGCTCAGAGGAGCACACCCATCATTTGG 1557

Qy     1328 TCAACTAATATTCCCCAGGGTTCAATAGATTAGCTCACTCTTTTGAAGGAAACGTGGTACTTAGTTA 1387
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1558 TCAACTAATATTCCCCAGGGTTCAATAGATTCACTCTTTTGAAGGAAACGTGGTACTTAGTTA 1617

Qy     1388 GGGTGGATGAAAAGCACAGAAGAAGAACTTACGCTCGCGGGGTTCCACTCCAAAATGATGACACTT 1447
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1618 GGGTGGATGAAAAGCACAGAAGAAGAACTTACGCTCGCGGGGTTCCACTCCAAAATGATGACACTT 1677

Qy     1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAACCGAGCATTATCCAAGCCCCTG 1507
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1678 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAACCGAGCATTATCCAAGCCCCTG 1737

Qy     1508 CCAAGAAAGTACCAAGACTCCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTGTCATTAGTA 1567
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1738 CCAAGAAAGTACCAAGACTCCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTA 1797

Qy     1568 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGGTGGGCATGGGGT 1627
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1798 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGGTGGGCATGGGGT 1857

Qy     1628 GGGGGTTATGGGGAAGAACAGTTGGGAGGAATGGGATATCTGGGGGGATAATTTAAAGGATTACATG 1687
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1858 GGGGGTATGGGAACAGTTGGGAGGAATGGGATATCTGGGGGATAATTTAAAGGATTACATG 1917

Qy     1688 TTATGTAAATTTTTATGTGACTGACATCGGAGCCTCGATGACTATCGTGTACTTGGGAAAG 1747
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1918 TTATGTAAATTTTTATGTGACTGACATCGGAGCCTCGATGACTATCGTGTACTTGGGAAAG 1977

Qy     1748 ATATGCAACACATGCCTATGATTTTTACAAGCCTGATATGGCTATCTCGAATATCCTATAGTAG 1797
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1838 ATATGCAACACATGCCTATGATTTTTACAAGCCTGATATGGCTATCTCGAATATCCTATAGTAG 897
```

```
RESULT 4
US-10-236-417-181
  Sequence 181, Application US/10236417
  Publication No. US20040048256A1
  GENERAL INFORMATION:
  APPLICANT: Agee et al.
  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
  FILE REFERENCE: 21402-442C
  CURRENT APPLICATION NUMBER: US/10/236,417
  CURRENT FILING DATE:  2003-01-06
  PRIOR APPLICATION NUMBER: US60/318,120
  PRIOR FILING DATE:  2001-09-01
  PRIOR APPLICATION NUMBER: US60/318,430
  PRIOR FILING DATE:  2001-09-10
  PRIOR APPLICATION NUMBER: US60/322,781
  PRIOR FILING DATE:  2001-09-17
  PRIOR APPLICATION NUMBER: US60/318,184
  PRIOR FILING DATE:  2001-09-07
  PRIOR APPLICATION NUMBER: US60/361,663
  PRIOR FILING DATE:  2002-03-05
  PRIOR APPLICATION NUMBER: US60/396,412
  PRIOR FILING DATE:  2002-07-17
  PRIOR APPLICATION NUMBER: US60/322,636
  PRIOR FILING DATE:  2001-09-17
  PRIOR APPLICATION NUMBER: US60/322,817
  PRIOR FILING DATE:  2001-09-17
  PRIOR APPLICATION NUMBER: US60/322,816
  PRIOR FILING DATE:  2001-09-17
  PRIOR APPLICATION NUMBER: US60/323,519
  PRIOR FILING DATE:  2001-09-19
  Remaining Prior Application data removed - See File Wrapper or PALM.
  NUMBER OF SEQ ID NOS: 341
  SOFTWARE: Custom
  SEQ ID NO 181
  LENGTH: 1650
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (22)..(1582)
  US-10-236-417-181

Query Match             68.4%;  Score 1370;  DB 13;  Length 1650;
Best Local Similarity   92.0%;
Matches 1509;  Conservative   0;  Mismatches   5;  Indels 126;  Gaps   1;
```

```
Qy         11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy        134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 TGGGAAATTGTTGCCCTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
```

```
Db         71 TTGGGAATTGTTGCCCTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 130

Qy        254 AAAAATATGATGGTGGTTGAGTACTCGGAAGTAATACCATTGGCTTGGGCCCAAATGTACT 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 AAAAATATGATGGTGGTTGAGTGCTCGGAAGTAATACCATTGGCTTGGGCCCAAATGTACT 190

Qy        314 GCTTCTGCACAGATAGGAAGAATTAACTCTCTTTTGCATGATGACTGTTGGAAATCTTA 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 GCTTCTGCACAGATAGGAAGAATTAACTCTCTTTTGCATGATGACTGTTGGAAATCTTA 250

Qy        374 TGGGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGGAAGTTGGAACAGAGCAA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 TGGGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGGAAGTTGGAACAGAGCAA 310

Qy        434 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGCTGTTTGAAGAGTTGGGA 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGCTGTTTGAAGAGTTGGGA 370

Qy        494 ATACAGATTATAGAAGGAAATGACACAACTAATGCATCGCTATGGGAGCGCACAGCTGTCT 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 ATACAGATTATAGAAGGAAATGACACAACTAATGCATCGCTATGGGAGCGCACAGCTGTCT 430

Qy        554 TCAATGCTGTTAACTGGGATTGAGTCCAGCTCCTTGGGAT------------------- 591
              ||||||||||||||||||||||||||||||||||||||||||
Db        431 TCAATGCTGTTAACTGGGATTGAGTCCAGCTCCTTGGGAT----------GGGCTCGTG 490

Qy        592 -------------------------------------------------------- 591
Db        491 CAGGAGATATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTGGAGCAG 550

Qy        592 ----------------------------------------------------- 591
Db        551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGGGGCTTCGTGTGGGCAC 610

Qy        608 ATATCGCAACATGCCTTATGATTTTTACAAGCCTGATTGTCTGAATATCCTATAGTAG 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        611 ATATCGCAACATGCCTTATGATTTTTACAAGCCTGATTGTCTGAATATCCTATAGTAG 670

Qy        668 ATGGAAAACTCTCAACAGTGCTACCTCAGTGCATTAGACCGGCTGCTACTCGTCTACT 727
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        671 ATGGAAAACTCTCAACAGTGCTACCTCAGTGCATTAGACCGGCTGCTACTCGTCTACT 730

Qy        728 GCAAAAAGATCCATGCCCAGTGGCAGAGAAGTGATGAAAGATTTTACCTTGAATG 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        731 GCAAAAAGATCCATGCCCAGTGGCAGAGAAGTGATGAAAGATTTTACCTTGAATG 790

Qy        788 ATTTTGGGCTTCACTGATCTTTCACTCACCATAGTGTAAACTGGTTCAGAAATCTTCTAGCTC 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 ATTTTGGGCTTCACTGATCTTTCACTCACCATAGTGTAAACTGGTTCAGAAATCTTCTAGCTC 850

Qy        848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAGAGATATAGTANTCTATAGTG 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTANTCTATAGTG 910

Qy        908 GCCTGGAAGCCTTTGGGGATTGTTAAATTAGGAAGACACCTACTTGATGAGGATGTCGGAGA 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGACACCTACTTGATGAGGATGTCGGAGA 970

Qy        968 AGGCATTATTGAAGGCTAGCTCTGAACTCTTCAGTCGAGAAAACAAAGGCATCTTTACTTG 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        971 AGGCATTATTGAAGGCTAGCTCTGAACTCTTCAGTCGAGAAAACAAAGGCATCTTTACTTG 1030

Qy       1028 TATCAAAATCAAAAATATGAAAATATGTACAATCTTCAGTATATGGTTCCCTGTCGCATCTGTTC 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 TATCAAAATCAAAAATATGAAAATATGTACAATCTTCAGTATATGGTTCCCTGTCGCATCTGTTC 1090

Qy       1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAGAATTGGAGTGTTTTTCTTATGGGTT 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAGAATTGGAGTGTTTTTCTTATGGGTT 1150

Qy       1148 CTGGGTTTGGGCTGCGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 CTGGGTTTGGGCTGCGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1210
```

```
Qy       1748 TCTCTTTGCTCTATTTGCTGACATGCTTCCTGTGTGGTCTCGGCCAATGCCAAAATGTACT 1807
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1978 TCTCTTTGCTCTATTTGCTGACATGCTTCCTGTGTGGTCTCGGCCAATGCCAAAATGTACT 2037

Qy       1808 CGAATGATGTTAAGGGCTTCTGTAAAAACTTCATACCTCGTTGTGTGGGTCTGGCCAATGCCAAAATGTATCGCATGATG 1867
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1808 CGAATGATGTTAAGGGCTTCTGTAAAAACTTCATACCTCGTTGTGTGGGTCTGGCCAATGCCAAAATGTATCGCATGATG 1867

Db       2038 CGAATGATGTTAAGGGCTTCTGTAAAAACTTCATACCTCGTTGTGTGGGTCTGGCCAATGCCAAAATGTATCGCATGATG 2097

Qy       1868 TTTGGTTTTTAAACATGGTATATAATGAATGTAATGTGTACTTCGTCAGAAGAAGAAGCAGAGGTAC 1927
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2098 TTTGGTTTTTAAACATGGTATATAATGAATGTAATGTGTACTTCGTCAGAAGAAGAAGCAGAGGTAC 1927

Qy       1928 TAATCTCCAATTAAAAAATTTTTTAATCATGTAAGAATTCGTGTACTTCGTCAGAAGAAGAAGCAGAGGTAG 2157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1928 TAATCTCCAATTAAAAAATTTTTTAATCATGTAAGAATTCGTGTACTTCGTCAGAAGAAGAAGCAGAGGTAG 1987

Qy       2158 TAATCTCCAATTAAAAAATTTTTTAATCATGTAAGAATTCGTGTACTTTGAACAACACAAAGATT 2217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2158 TAATCTCCAATTAAAAAATTTTTTAATCATGTAAGAATTCGTGTACTTTGAACAACACAAAGATT 2217

Qy       1988 AAAAAAA 1994
              |||||||
Db       2218 ACAGAAA 2224
```

```

```
Qy   1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTT 1447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGATGAAAAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTT 1450

Qy   1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1507
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTG 1510

Qy   1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTA 1567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTA 1570

Qy   1568 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGCATGGGGT 1627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGCATGGGGT 1630

Qy   1628 GGGGGTATGGGAACAGTTGG 1647
              ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650

RESULT 5
US-10-236-417-183
; Sequence 183, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 183
;   LENGTH: 1650
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22)..(1582)
US-10-236-417-183

  Query Match   68.4%; Score 1370; DB 13; Length 1650;
  Best Local Similarity   92.0%; Pred. No. 0;
  Matches 1509; Conservative   0; Mismatches   5; Indels 126; Gaps   1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
              | || |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 TTGGGATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   131 AAAAATATGATGGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 373
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTA 250

Qy   374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA 310

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 493
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCT 553
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 ATACAGATATAGAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCT 430

Qy   554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT---------------------- 591
              ||||||||||||||||||||||||||||||||||||||
Db   431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490

Qy   592 -- 591

Db   491 CAGGAGATATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAG 550

Qy   592 --GGGCTTCGTGGGACAC 607
                                                          ||||||||||||||||
Db   551 TAGCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACAC 610

Qy   608 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670

Qy   668 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727
              |||| |||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   671 ATGGGAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACT 730

Qy   728 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATG 787
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATG 790

Qy   788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 850

Qy   848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTG 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTG 910

Qy   908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGA 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGA 970

Qy   968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTG 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db    971 AGGCCATTTATGAGGCTAGCTCTGAAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTG 1030
Qy    134 GCTCTTTTCACCTATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193

Db   1028 TATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATAGGTTCCCTTGCATCTGTTC 1087
Qy     11 GCTCTTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Db   1031 TATCAAAATCAAAATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC 1090
Qy    194 TGGGAAATTGTTGCCCTTGAGATCTATATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 253

Db   1088 TAGCCACAGTACTCACCTCAGCATTAGCAGGGGAGGAGAGAGAATTGGAGTGTTGTTTCTTATGGTT 1147
Qy     71 TTGGGAATTGTTGCCCCTTGAGATCTATATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 130

Db   1091 TAGCCACAGTACTCACCTCAGCAATTAGCACAGAGAGAATTGGAGTGTTTTCTTATGGTT 1150
Qy    253 TGGGAAATTGTTGCCCTTGAGATCTATTTCCTTCTCTATATGTTGATCAAGCACAGTTG 312

Db   1148 CTGGTTTCGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1207
Qy    194 TGGGAAATTGTTGCCCTTGAGATCTATATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 253

Db   1151 CTGGTTTCGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTG 1210
Qy     71 TTGGGATTGTTGCCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 130

Db   1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
Qy    254 AAAAATATGATGGTGTTAGATGCTGGAAAGTAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313

Db   1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1270
Qy    131 AAAAATATGATGGTGTTAGATGCTGGAAAGTAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 190

Db   1268 GTGTGGCACCAGATGTCTTCGCTGGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
Qy    314 GCTTCTGCACGATAGAGATAGAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 373

Db   1271 GTGTGGCACCAGATGTCTTCGCTGGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1330
Qy    191 GCTTCTGCACGATAGAGATAGAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 250

Db   1328 TCAACTATATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAACGTGGTACTTAGTTA 1387
Qy    374 TGGGAGAAATAAACCTTTTTCCTATGATAAGAGAAGATATTAACTTTGGCATGACTGTGGT 433

Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAACGTGGTACTTAGTTA 1390
Qy    251 TGGGAGAAATAAACCTTTTTCCTATGATTGCAATGGGCGGCGTGGGAACTGGAGTTGGAAC 310

Db   1388 GGGTTGGATGAGAAAAGCACAGAAGAACTTAGCGTCTCGGGGTCCCAGTCCCAAAATGATGACACTT 1447
Qy    434 TCATCGACAAAATCAAAAGTTCGTGAAGACTAAATTTCAATGCAAGCTGTTTGAAGAGTCTGGGA 493

Db   1391 GGGTTGGATGAGAAAAGCACAGAAGAACTTAGCGTCTCGGGGTCCCAGTCCCAAATGATGACACTT 1450
Qy    311 TCATCGACAAATCAAAAGTTCGTGAAGACTAAATTTCAATGCAAGCTGTTTGAAGAGTCTGGGA 370

Db   1448 TGGATGAAGGAGTAGGGACTTGTCATTCAAAACATAGCAACTACACTAATTCCAAGCCCTG 1507
Qy    494 ATACAGATATAGAAGGAAATCGACACAACTAATGCAATGCTAATGGAGGCACACAGCTGCTGTCT 553

Db   1451 TGGATGAAGGAGTAGGGACTTGTCATTCAAAACATAGCAACTACACTAATTCCAAGCCCTG 1510
Qy    371 ATACAGATATAGAAGGAAATCGACAACTAATGCAATGCTAATGGAGGCACACAGCTGCTGTCT 430

Db   1508 CCAAGAAAGTACCAGAGACTCCCTGCCACAGCGAGAGGAACCTGAAGCACCTGATTGCTTAGTA 1567
Qy    314 GCTTCTGCACGATAGAGATAGAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 373

Db   1511 CCAAGAAAGTACCAGAGACTCCCTGCCACAGCGAGAGGAACCTGAAGCACCTGATTGCTTAGTA 1570
Qy    494 ATACAGATATAGAAGGAAATCGACAACTAATGCAATGCTAATGGAGGCACACAGCTGCTGTCT 553

Db   1568 ATGGGGAACAATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT 1627
Qy    554 TCAATGCTGTTTACTGGATTTGAGTCCAGCTCTTGGGAT------------------ 591

Db   1571 ATGGGGAACAATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT 1630
Qy    592 --------------------------------------------------- 591

Db   1628 GGGGGTATGGGGAACAGTTGG 1647
Qy    592 ----------------------------------------------------- 591

Db   1631 GGGGGTATGGGGAACAGTTGG 1650
Qy    491 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGGTTGGAGCAG 550
```

Query Match 68.4%;   Score 1370;   DB 13;   Length 1650;
Best Local Similarity 92.0%;   Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

```
Qy    134 GCTCTTTTCACCTATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
Db     11 GCTCTTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy    194 TGGGAAATTGTTGCCCTTGAGATCTATATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 253
Db     71 TTGGGAATTGTTGCCCCTTGAGATCTATTTCCTTCTCAATATGTTGATCAAGCACAGTTG 130

Qy    254 AAAAATATGATGGTGTTAGATGCTGGAAAGTAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313
Db    131 AAAAATATGATGGTGTTAGATGCTGGAAAGTAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 190

Qy    314 GCTTCTGCACGATAGAGATAGAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 373
Db    191 GCTTCTGCACGATAGAGATAGAGATATTAACTCTCTTGCATGACTGTGGTTCAGAATCTTA 250

Qy    374 TGGGAGAAATAAACCTTTTTCCTATGATTGCAATGGGCGGCGTGGGAACTGGAGTTGGAAC 433
Db    251 TGGGAGAAATAAACCTTTTTCCTATGATTGCAATGGGCGGCGTGGGAACTGGAGTTGGAAC 310

Qy    434 TCATCGACAAATCAAAAGTTCGTGAAGACTAAATTTCAATGCAAGCTGTTTGAAGAGTCTGGGA 493
Db    311 TCATCGACAAATCAAAAGTTCGTGAAGACTAAATTTCAATGCAAGCTGTTTGAAGAGTCTGGGA 370

Qy    494 ATACAGATATAGAAGGAAATCGACAACTAATGCAATGCTAATGGAGGCACACAGCTGCTGTCT 553
Db    371 ATACAGATATAGAAGGAAATCGACAACTAATGCAATGCTAATGGAGGCACACAGCTGCTGTCT 430

Qy    554 TCAATGCTGTTTACTGGATTTGAGTCCAGCTCTTGGGAT---------------- 591
Db    431 TCAATGCTGTTTAACTGGATTTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490

Qy    592 ---------------------------------------------- 591
Db    491 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGGTTGGAGCAG 550

Qy    592 ----------------------------------------GGGCTTCGTGGGCAC 607
Db    551 TAGCTCTGCTAAATGGGGCCAAATGCTCCTTAATTTTGAACGAGGGCTTCGTGGGCAC 610

Qy    608 ATATGCCAACATGCCTATGATTGATTTTTACAACGCCTGATAATATCCTATAGTAG 667
Db    611 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATCCTATAGTAG 670

Qy    668 ATGGAAAACTCTTCCATACAGTGCTRACCTCAGTGCATTAGACCGCGCTTACTCTCGTCTACT 727
Db    671 ATGGGAAACTCTTCCATACAGTGCTRACCTCAGTGCATTAGACCGCGCTTACTCTCGTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAGAAAGAGGGAAATGATAAAAGATTTTTACCTTGAAATG 787
Db    731 GCAAAAAGATCCATGCCCAGTGGCAGAGAAAGAGGGAAATGATAAAAGATTTTTACCTTGAAATG 790

Qy    788 ATTTTGGCTTCATGATCTTCACTCACCATTATTGTAAACTGGTTCAGAAATCTCTAGCTC 847
```

```
RESULT 7
US-10-236-417-189
Sequence 189, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
  APPLICANT: Agee et al.
  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
  FILE REFERENCE: 21402-442C
  CURRENT APPLICATION NUMBER: US/10/236,417
  CURRENT FILING DATE: 2003-01-06
  PRIOR APPLICATION NUMBER: US60/318,120
  PRIOR FILING DATE: 2001-09-01
  PRIOR APPLICATION NUMBER: US60/318,430
  PRIOR FILING DATE: 2001-09-10
  PRIOR APPLICATION NUMBER: US60/322,781
  PRIOR FILING DATE: 2001-09-17
  PRIOR APPLICATION NUMBER: US60/318,184
  PRIOR FILING DATE: 2001-09-07
  PRIOR APPLICATION NUMBER: US60/361,663
  PRIOR FILING DATE: 2002-03-05
  PRIOR APPLICATION NUMBER: US60/396,412
  PRIOR FILING DATE: 2002-07-17
  PRIOR APPLICATION NUMBER: US60/322,636
  PRIOR FILING DATE: 2001-09-17
  PRIOR APPLICATION NUMBER: US60/322,817
  PRIOR FILING DATE: 2001-09-17
  PRIOR APPLICATION NUMBER: US60/322,816
  PRIOR FILING DATE: 2001-09-17
  PRIOR APPLICATION NUMBER: US60/323,519
  PRIOR FILING DATE: 2001-09-19
  Remaining Prior Application data removed - See File Wrapper or PALM.
  NUMBER OF SEQ ID NOS: 341
  SOFTWARE: Custom
SEQ ID NO 189
  LENGTH: 1650
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (22)..(1582)
US-10-236-417-189

Query Match     68.4%;  Score 1370;  DB 13;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 0;
Matches 1509;  Conservative   0;  Mismatches 5;  Indels 126;  Gaps 1;
```

```
Qy  134 GCTCTTTCACCATGCCTGGATCACTTCCCTTGAAGTGCAGAAGAAGATG 193
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 GCTCTTTCACCATGCCTGGATCACTTCCCTTGAAGTGCAGAAGAAGATG  70

Qy  194 TGGGAATTGTTGCCCTTGAGATCTTATTTCTTCCTTTCTCAAATATGTTGATCAGACAGAGTTGG 253
         || |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db   71 TGGGAATTGTTGCCCTTGAGATCTTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG    130

Qy  254 AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG 313
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  131 AAAAATATGATGGTGTTAGATGCTGGAAGTATACCATTGGCTTGGGCCAGGCCAAGATGG    190

Qy  314 GCTTCTGCACAGATAGAGGAGGATATTAACTCTCCTTTGCATCACTGTGGTTCAGAATCTTA 373
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  191 GCTTCTGCACAGATAGAGAAGGATATTAACTCTCCTTTGCATGATGGTTCAGAATCTTA    250

Qy  374 TGGGAGAGAAATAAACCTTTTCCTATGATTGCATTGGGCCGCGGCTGGGAGTTGGAAACAGAACAA 433
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  251 TGGGAGAGAAATAAACCTTTTCCTATGATTGCATTGGGCCGCGGCTGGGAGTTGGAAACAGAACAA    310

Qy  434 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGCTGTTTTGAAGAGTCTGGGA 493
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTTGAAGAGTCTGGGA    370

Qy  494 ATACAGATAATAGAAGGAAATCGACACAACTAATTGCATGCTATGGAGGGCACAGCTGCTGCT 553
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  371 ATACAGATAATAGAAGGAAATCGACACAACTAATTGCATGCTATGGAGGGCACAGCTGCTGCT    430

Qy  554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTTAGTTG 591
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTTAGTTG    490

Qy  592 -------------------------------------------------------------- 591
Db  491 CAGGAGATAATTGCTGTATATGCCACAGGAAATGCTACGACCTTACAGGTGGAGTTGGGAGCAG    550

Qy  592 -------------------------------------------------------------- 591
Db  551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTTGAACGAGAGGGCTTCGTGTGGGACAC    610
```

```
Qy  791 ATTTTGGCTTCATGACTACTTTCACTCACCACCATATTGTAAACTGGTTCAGAAAATCTAGCTC 850
Db  ...

Qy  848 GGATCGTTTGCTGAATGACTTCCTTAATGACCACGAATAGAGATAAAATAGTATCTATAGTG 907
Db  851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAAAATAGTATCTATAGTG 910

Qy  908 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGAAGACACCTTTGGAGATGTGGAGA 967
Db  911 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGAAGACACCTTTGGAGATGTGGAGA 970

Qy  968 AGGCATTTATGAGGCTAGCTCTTGAACTTCAGTCGTCAGAAAAGACAAAGGCATCTTTACTTG 1027
Db  971 AGGCATTTATGAGGCTAGCTCTTGAACTTCAGTCGTCAGAAAAGACAAAGGCATCTTTACTTG 1030

Qy 1028 TATCAAAATCAAAATGGAAAATAATGTACACATCTTCAGTATATATGGTTCCCTTGCATCTGTTC 1087
Db 1031 TATCAAAATCAAAATGGAAAATAATGTACACATCTTCAGTATATATGGTTCCCTTGCATCTGTTC 1090

Qy 1088 TAGCACAGTACTCACCTCAGCAATTAGCACAGGGAAGAGAATTGGAGTGTTTTTCTTATGGTT 1147
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCACAGGGAAGAGAATTGGAGTGTTTTTCTTATGGTT 1150

Qy 1148 CTGGTTTGGCTGCCACTCTGTACTCTCCTTAAAGTCACAAGATGTCACACCGGGGTCTG 1207
Db 1151 CTGGTTTGGCTGCCACTCTGTACTCTCCTTAAAGTCACAAGATGTCACACCGGGGTCTG 1210

Qy 1208 CTCTTGATAAAATAAACACGACAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
Db 1211 CTCTTGATAAAATAACACGACAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 1270

Qy 1268 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTGG 1327
Db 1271 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTGG 1330

Qy 1328 TCAACTATTATTCCCAGGGTTCAATAGAATTCACTCTCTTTGAAGGAAACGTGGTACTTAGTTA 1387
Db 1331 TCAACTATTATTCCCAGGGTTCAATAGAATTCACTCTCTTTGAAGGAAACGTGGTACTTAGTTA 1390

Qy 1388 GGGTGGATGAAAAGCACAGAGAAGAACTTACGTTCGGGCGTCCCACTCCAAATGATGACACTT 1447
Db 1391 GGGTGGATGAAAAGCACAGAGAAGAACTTACGTTCGGGCGTCCCACTCCAAATGATGACACTT 1450

Qy 1448 TGGATGAAGGAGTAGGACTTGTGTCATTCAAACATAGCACATAGACCCCTG 1507
Db 1451 TGGATGAAGGAGTAGGACTTGTGTCATTCAAACATAGCACATAGACCCCTG 1510

Qy 1508 CCAAGAAAGTACCAAGAGATTCCCTGCCACAGCAGCAGAGAAGCTGAAGCAGTGTCATTAGTA 1567
Db 1511 CCAAGAAAGTACCAAGAGATTCCCTGCCACAGCAGCAGAGAAGCTGAAGCAGTGTCATTAGTA 1570

Qy 1568 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGAAACTTCAGGGTGGGTGGGCATGGGGT 1627
Db 1571 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGAAACTTCAGGGTGGGTGGGCATGGGGT 1630

Qy 1628 GGGGGTATGGGAACAGTTGG 1647
Db 1631 GGGGGTATGGGAACAGTTGG 1650
```

RESULT 8
US-10-236-417-193
Sequence 193, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US/10/236,417
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 193
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-193

Query Match   68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

Qy   134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAGACTTGCTGGCCAAAAGATG 193
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAGACTTGCTGGCCAAAAGATG 70

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG 253
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 TTGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAAATATGATGGGTGTAGATGGTCTGGAAAGTATACCATTGGCTTGGCCTCAGGCATG 313
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAAATATGATGGGTGTAGATGGTCTGGAAAGTATACCATTGGCTTGGCCTCAGGCATG 190

Qy   314 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTCGGTTCAGATCTTA 373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTCGGTTCAGATCTTA 250

Qy   374 TGGAGAGAAATAACCTTTCCTTATAGATTGCATTGGGCGGGGCTGGAAGTTGGAACAGTAA 433
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGAGAAATAACCTTTCCTTATAGATTGCATTGGGCGGGGCTGGAAGTTGGAACAGTAA 310

Qy   434 TCATCGACAAATCAAAGTCTGTGAGAGACTAAATTTGATGCAGCTCGTTTGAAGAGTCTGGGA 493
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTTGATGCAGCTCGTTTGAAGAGTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAATCGACACAACTAAATGCATGCTAATGGAGGCACAAGCTGCTCT 553
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 ATACAGATATAGAAGGAATCGACACAACTAAATGCATGCTAATGGAGGCACAAGCTGCTCT 430

Qy   608 ATATGCAACACATGCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 ATATGCAACACATGCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670

Qy   668 ATGGAAAAACTCTCCATACACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 ATGGAAAAACTCTCCATACACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT 730

Qy   728 GCAAAAAGATCCATGCCCAGTGCCAGAAAAGAGGGAAAATGATAAAGAGATTTTACCTTGAATG 787
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 GCAAAAAGATCCATGCCCAGTGCCAGAAAAGAGGGAAAATGATAAAGAGATTTTACCTTGAATG 790

Qy   788 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTTAAACTGGTTCAGAGAAATCTCTAGCTC 847
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTTAAACTGGTTCAGAGAAATCTCTAGCTC 850

Qy   848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAGAAAAATAGTTATAGTATAGTG 907
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATAGAAAAATAGTTATAGTATAGTG 910

Qy   908 GCCTGGGAAGCCTTTGGGGATGTTAAATTAGAAGAGCACCTACTTTGATGAGAGATGTGGAGA 967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 GCCTGGGAAGCCTTTGGGGATTGTTAAATTAGAAGAGCACCTACTTTGATAGAGATGTGGAGA 970

Qy   968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGGCATCTTTACTTG 1027
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAGGCATCTTTACTTG 1030

Qy 1028 TATCAAAATCAAAATGGAAAATATAGTACACATCTTCAGTATATATGGATGTGGAGA 1087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1031 TATCAAAATCAAAATGGAAAATATGTACACATCTTCAGTATATATGGTTCCATCTGTTC 1090

Qy 1088 TAGGCACAGTACTCACCTCAGCAATTAGCAGGAAGAGAGAGAATTGGAGTGTTTTCTTATGGTT 1147
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGAAGAGAGAGAATTGGAGTGTTTTCTTATGGTT 1150

Qy 1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATCTACACCGGGGTCTG 1207
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1151 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATCTACACCGGGGTCTG 1210

Qy 1208 CTCTTTGATAAAAATAAACAGCAAAGTTTATGTGATCTTAAAATCAAGCCTTGATTCAAGAAACTG 1267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1211 CTCTTTGATAAAAATAAACAGCAAAGTTTATGTGATCTTAAAATCAAGCCTTGATTCAAGAAACTG 1270

Qy 1268 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGGACTCAGAGAGAATCTTGG 1327
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1271 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGGACACCCATCATTTGG 1330

Qy 1328 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTCTTGAAGGAAACGTGGTACTTAGTTA 1387
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1331 TCAACTATATTCCCAGGGTTCAATAGATTCACTCTCTTGAAGGAAACGTGGTACTTAGTTA 1390

Qy 1388 GGGTGGATGGAAAAGCACAGAGAGAACTTACGGCTCGGCGTCCCACTCCAAATGATGATGACACTT 1447
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1391 GGGTGGATGGAAAAGCACAGAGAGAACTTACGGCTCGGCGTCCCACTCCAAATGATGATGACACTT 1450

Qy 1448 TGGATGGAGGAGTAGGACTTGTTGCATTCAAAACATAGCAACTAGCACATATTCCAAGCCCTG 1507
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1451 TGGATGGAGGAGTAGGACTTGTTGCATTCAAAACATAGCAACTAGCACATATTCCAAGCCCTG 1510

Qy 1508 CCAAGAAAGTACCAAGAGACTCCCGCCACAGCAGCAGCACTGTCATTAGTTA 1567
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1511 CCAAGAAAGTACCAAGAGACTCCCTCCCACAGCAGCAGCACTGTCATTAGTTA 1570

Qy 1568 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGCATGGGGGT 1627
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1571 ATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGCATGGGGGT 1630

Qy 1628 GGGGGTATGGGAACAGTTGG 1647
         ||||||||||||||||||||
Db 1631 GGGGGTATGGGAACAGTTGG 1650

```
RESULT 9
US-10-236-417-197
; Sequence 197, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-442C
;  CURRENT APPLICATION NUMBER: US/10/236,417
;  CURRENT FILING DATE: 2003-01-06
;  PRIOR APPLICATION NUMBER: US60/318,120
;  PRIOR FILING DATE: 2001-09-01
;  PRIOR APPLICATION NUMBER: US60/318,430
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: US60/322,781
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/318,184
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US60/361,663
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US60/396,412
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: US60/322,636
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,817
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,816
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/323,519
;  PRIOR FILING DATE: 2001-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom
; SEQ ID NO 197
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (22)..(1582)
US-10-236-417-197

Query Match             68.4%;  Score 1370;  DB 13;  Length 1650;
Best Local Similarity   92.0%;
Matches 1509;  Conservative    0;  Mismatches    0;  Indels  126;  Gaps   1;
```

```
Qy    134 GCTCTTTCACCATGCCTGGATCACTCCTTTGGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
          ||||||||||||||||||||||||| ||||||| ||||||||||||||||||| |||||||
Db     11 GCTCTTTCACCATGCCTGGATCACATCACTTCCTTTGAATGCAGAAGCTTGCTGCCAAAAGATG 70

Qy    194 TGGGAATTGGTTGCCCTTGAGATCTAATTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 253
          | ||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db     71 TTGGGAATTGGTTGCCCCTTGAGATCTTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 130

Qy    254 AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGAGTGG 313
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    131 AAAAAATATGATGGTGTAGATGCTGGAAGTATACCATTGGCTTGGGCCAGGCCAAGAGTGG 190

Qy    314 GCTTCTGCACAGATAGGAGGAGATATTAACTCTCTTTGCATGACTCGTGGGTTCAGAAATCTTA 373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    ...
```

```
Qy    491 CAGGAGGATAATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGGAGCAG 550
Db    551 TAGCTCTTGCTTAATTGGGCCAAATGCTCCTTTAATTTTTGAACGACGGGCTTCGTGGGGACAC 610

Qy    554 TCAATGCTCTGTTAACTGGATTGAGTCTCAGCTCTTGGGAT------------------- 591
          ||||| ||||||||||||||||||||| ||||||||||||
Db    431 TCAATCGTCTGTTAACTGGATTGAGTCCAGCTCTTGGGGATGGACGGTATGCCCTGGTAGTTG 490

Qy    592 ------------GGGGCTTCGTTGGGACAC 607
                      ||||||||| |||||||
Db    551 TAGCTCTTGCTTAATTGGGCCAAATGCTCCTTTAATTTTTGAACGACGGGCTTCGTGGGGACAC 610

Qy    608 ATATGCCAACATGCCTATATATTTTTACAAGCCTGATATGTCTATCTGAATATCCTATAGTAG 667
          ||||||| ||||||||||| |||||||||   |||||| |||| ||||||||||||||||||
Db    611 ATATGCAAACATGCCTATGATTTTTACAGACGCTATCTGAAATATCCTATAGTAG 670

Qy    668 ATGGAAAAACTCTCCATACATGCTACCTCACGTGCCATTACAGACCGCTGCTTCCTCTGTCTACT 727
          |||| ||||||||| |||||| ||||||||| ||| |||| ||||||||| |||| ||||||||
Db    671 ATGGGAAAACTCCCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCATTCCTGTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGCATTAAAGATTTTACCTTGAATG 787
          ||||||||||| |||||||||||| ||||||||||||||||||| ||||||||| |||||||
Db    731 GCAAAAAGATCCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATG 790

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACGTGGTTTCAGAAATCTCTAGCTC 847
          ||||||||||||| ||||||||||||||||||| |||||   ||||||||||||||||||||
Db    791 ATTTTGGCTTCATGAGTCTTTCACTCACCCATATTGTAAACTGGTTTCAGAAATCTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGAACTTCCTTAATGACCAGAGATAAAAATAGTACTTCATAGTG 907
          |||||||||||||||| ||||||||||| |||||| |||||| ||| ||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAGATAAAAATAGTACTTCATAGTG 910

Qy    908 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGAAGAACACCTACTTCAGTCAGTGAGAGTGAGA 967
          |||||||||||||||||||| |||||| ||||||||||||||| ||||||||| |||||||||
Db    911 GCCTGGAAGCCCTTTGGGGATGTTAAATTAGAAGAACACCTACTTCAGTCAGTGAGAGTGAGA 970

Qy    968 AGGCATTTATGAAGGCTAGCTCTGACTTTCAGTCAGAAAAACAAAAGGCATCTTTACTTG 1027
          |||||||||| ||||| |||| |||||||||||||||||||| ||||||||||||||||||
Db    971 AGGCATTTATGAAGGCTAGCTCTGAACTTTCAGTCAGTAAAAACAAAAGGCATCTTTACTTG 1030

Qy   1028 TATCAAAATCAAAAATGGAAAATAATGTACACATCTTCACTAATATAGTGGTTCCCTTGCAATCTGTTTC 1087
          |||||||| |||||| ||||| ||||| |||||||||||||| |||||||| ||||| |||||||||||||
Db   1031 TATCAAAATCAAAAATGGAAAATAATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCACCTCAGCAATTAGCAGGGAAGAGAAGTTGGACTGTTTTCTTATGGTT 1147
          |||||||||||||||||||||||||||| ||||||||| ||||| |||||| |||||||||||||
Db   1091 TAGCACAGTACTCACCTCACCTCAGCAATTAGCAGGGAAGAGAATTGGACTGTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTTAAGTCACACAGAGTCTCACCGGGGGTCTG 1207
          || ||||||| ||| ||||||||||||| ||||||| ||||||||||||| |||||||||
Db   1151 CTGGTTTGGGCTGCCACTCTGTACTCTCTTTAAAGTCACAAGAGATGCTACACCGGGGTCTG 1210

Qy   1208 CTCTCTGATAAAATAACACGCAAGTTTATGTGATCTTAAATCAAGCTTGATTCAAGAACTG 1267
          ||||| |||| ||||| || |||||||| || |||| |||| |||||| |||||||||||||
Db   1211 CTCTTCGATAAAATAACAGCAAGTTCTGTGATCTTAAATCAAGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGCACCCATCATTTGG 1327
          ||||||||||||| ||||||||||| |||||||| |||||||| |||||| ||||||||
Db   1271 GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGCACCCATCATTTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAGCGTGGTACTTAGTTA 1387
          ||||| |||| |||||||||||||||| ||||||||||||||||| ||||||| ||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAAGCGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAACACAGAGAAGAACTTTACGCTGCGGGGGTTCCCACTCCCAAATGAATGAGACACTT 1447
          |||||||| ||||| ||||||||| |||||||||||||||||||||||||||| |||||||||||||||||
Db   1391 GGGTGGATGAAAAACACAGAGAAGAACTTACGCTGCGGGGTTCCCACTCCCAAATGATGAACATT 1450

Qy   1448 TGGATGAAGGAGTAGGACTTGTTGCATTCAAAACATAGCAACTTAATTCCAAGCCCTG 1507
          |||||||||||| |||||||||||||||||||||||||||| || |||||||||||||
Db   1451 TGGATGAAGGAGTAGGACTTGTTGCATTCAAAACATAGCAACTTATTCCAAGCCCTG 1510
```

```
Qy   1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTGTCATTAGTA 1567
          ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTGTCATTAGTA 1570

Qy   1568 ATGGGGGAACATTAAGATACTCTGTGAGGTGCAAAGACTTCAGGGTGGGGTGGGGGTGGGGGT 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGGAACATTAAGATACTCTGTGAGGTGCAAAGACTTCAGGGTGGGGTGGGGGTGGGGGT 1630

Qy   1628 GGGGGTATGGGAACAGTTGG 1647
          ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650
```

```
Qy    191 GCTTCTCGACAGATAGAGAAGATATTAACTCTCTTTGCATGACTGTGTGTTCAGAATCTTA 250
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA 433

Qy    251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCCTGGAAGTTGGAACAGAGACAA 310
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCCAGTCTGTTGAAGAGTCTGGGA 493

Qy    311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATCCAGTCTGTTGAAGAGTCTGGGA 370
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    494 ATACAGATATAGAAGGAATCGACAACAACTAATCGCATAGCTGAGGAGGCAACAGCTGCTGCT 553

Qy    371 ATACAGATATAGAAGGAATCGACAACAACTAATCGCATAGCTGAGGAGGCAACAGCTGCTGCT 430
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 553

Qy    431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490
              ||||||||||||||||||||||||||||||||||||||||||||||-----------
Db    554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT----------- 591

Qy    491 CAGGAGATATTGCTGTATATGCCACAGGAAATGCTAGACCTACACAGGTGGAGTGGCAG 550
                                                      ------
Db    592 ----------------------------------------- 591

Qy    551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTTGAACGAGGCTTCGTGGGGACAC 610
                                           ------GGGGCTTCGTGGGGACAC
Db    592 ----------------------------------GGGGCTTCGTGGGGACAC 607

Qy    608 ATATGCAACACATGCCTATGATTGATTTTTTACAAGCCTGATATGCTATCTCTATAGTAG 667
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACATGCCTATGATTTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 670

Qy    668 ATGGAAAACTCTCCATACAGTGCTACCTCGAGTGCATTAGACCCGCTGCTACTCTGTCTACT 727
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCCGCTCTATTCTGTCTCACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAAAAGAGGGAAAATGATAAAGATTTTACCTTGAATG 787
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCATGCCCAGTGGCAGAAAAGAGGGAAAATGATAAAGATTTTACCTTGAATG 790

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATGTTGTCAGAAAATCTTCTAGCTC 847
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTGGCTTCATGATCTTTCACTCACCATATGTTGTCAGAAAATCTTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGTAAAAATAAGTATTCTATAGTG 907
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGTAAAAATAAGTATTCTATAGTG 910

Qy    908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGACACCTACTTTGGTAAAATAGTGGAGA 967
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGACACCTACTTTGGTAAAATAGTGGAGA 970

Qy    968 AGGCATTTATGAGAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCCATCTTTACTTG 1027
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAGAGGCTAGCTCTGAACTTCTCAGTCAGAAAACAAAGGCCATCTTTACTTG 1030

Qy   1028 TATCAAATCAAAATGGAAATATGTACACATCTCTCAGTATATGGTTCCCTTGCATCTGTTC 1087
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAATCAAAATGGAAATATGTACACATCTCTCAGTATATGGTTCCCTTGCATCTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCAGGGGAAGGAGGAATTGGAGTGTTTTCTTATGGTT 1147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCAGGGGAAGGAGGAATTGGAGTGTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGGTCTG 1207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGGTCTG 1210

Qy   1208 CTCTTGATAAAATAACAGCAAGTTTATTTGATGTATCTTAAATCAACAAGGCTTGATTCAAGAACTG 1267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATTTGATGTATCTTAAATCAACAAGGCTTGATTCAAGAACTG 1270

Qy   1268 GTGTGGGCACCAGGATGTCTTCGCTGGAAAAACATGGAAGCTCAGAGGACGAGGAGGGACACCCATCACATTTGG 1327
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGGCACCAGGATGTCTTCGCTGGAAAAACATGGAAGCTCAGAGGACGAGGAGGGACACCCATCACATTTGG 1330

Qy   1328 TCAACTATAATTCCCCAGGGGTTCAATTAGAATTCACTCTTTGAAGGAACGGTGGTACTTAGTTA 1387
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATAATTCCCCAGGGGTTCAATTAGAATTCACTCTTTGAAGGAACGGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGTGAAAAGCACAGAAGAACTTACGCTCGGGGTCCCACTCCAAATGATGACACTT 1447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGTGAAAAGCACAGAAGAACTTACGCTCGGGGTCCCACTCCAAATGATGACACTT 1450

Qy   1448 TGGATGAAGGAGTAGGACTTGTGCATTCAAAACATAGCACAACTGAGCATATTCCTCCTG 1507
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTAGGACTTGTGCATTCAAAACATAGCACAACTGAGCATATTCCTCCTG 1510

Qy   1508 CCAAGAAAAGTACCAAGACTCCCTCGCCACAGCAGCAGAAGCAGCGGAAGACCTGAGCAGCTG 1567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAAGTACCAAGACTCCCTCGCCACAGCAGCAGAAGCAGCGGAAGACCTGAGCAGCTG 1570

Qy   1568 ATGGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGGTGGGGGTGGGGCATGGGGGT 1627
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGGTGGGGGTGGGGCATGGGGGT 1630

Qy   1628 GGGGGTATGGGGAACAGTTGG 1647
              |||||||||||||||||||||
Db   1631 GGGGGTATGGGGAACAGTTGG 1650

Qy    134 GCTCTTCACCATGCCTGGATCACTTCCTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
```

RESULT 10
US-10-236-417-201
; Sequence 201, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; FILE REFERENCE: 21402-442C
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 201
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-201

Query Match 68.4%; Score 1370; DB 13; Length 1650;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1509; Conservative 0; Mismatches 5; Indels 126; Gaps 1;

```
Db    11  GCTTCTTCACCATGCCTGGGATCACTCACTTGCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG   70
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   194  TGGGAAGTTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG     253

Db    71  TTGGGATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG      130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   254  AAAAAATATGATGGTGTTAGAGATGCTGGAAAGTATACCATTGGCTTGGCCAAGATGG       313

Db   131  AAAAAATATGATGGTGTTAGAGATGCTGGAAAGTATACCATTGGCTTGGCCAAGATGG       190
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   314  GCTTCTTCCACAGATAGAGAAGAGATATTAACTCTCTCTTCTTGGCATGACTGTGGTTCAGATCTTA 373

Db   191  GCTTCTTCCACAGATAGAGAAGAGATATTAACTCTCTCTTCTTGGCATGACTGTGGTTCAGATCTTA  250
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   374  TGGAGAGAAAATAAACCTTTCCTATATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA     433

Db   251  TGGAGAGAAAATAAACCTTTCCTATATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAA     310
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   434  TCATCGACAAATCAAAGTCTCTGTGAAGACTAATTTGATCCAGCTGTTTGAACAGAGTCTTGGGA   493

Db   311  TCATCGACAAATCAAAGTCTCTGTGAAGACTAATTTGATCCAGCTGTTTGAACAGAGTCTTGGGA   370
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   494  ATACAGATATAGAGGAATCGACACACACTAATGCATCTATGGGAGGCACAGCTGCTGTGTCT     553

Db   371  ATACAGATATAGAGGAATCGACACACACTAATGCATCTATGGGAGGCACAGCTGCTGTGTCT     430
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   554  TCAAATGCTGTTTAACTGGATTGAGTCCAGCTCTCAGCTCTTGGGAT---------        591

Db   431  TCAAATGCTGTTTAACTGGATTGAGTCCAGCTCTCAGCTCTTGGGATGGACCGGTAGTGCCCTGGTAGTTG  490
          |||||||||||||||||||||||||||||||||||||||
Qy   592  -----------------------------------GGGGCTTCGTGGGAGGCTTCGTGGGACAC        607

Db   491  CAGGAGATAATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGGTGGGAGTTGGAGCAG        550
                                              |||||||||||||||||||||||
Qy   608  ATATGCAACATGCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG       667

Db   551  TAGCTCTGCTAATTGGGGCCAAAATGCTTCCTTTAATTTTGAACGAGGCGCTTCGTGGGACAC       610
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   592  ---------GGGGCTTCGTGGGACAC        607

Db   611  ATATGCAACATGCCTATGATTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG       670
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   668  ATGGAAAAACTCTCCATACAGTGCTACCTCACTGCACCTGCTACTCACTCGTGTCTACT       727

Db   671  ATGGAAAAACTCTCCATACAGTGCTACCTCACTGCACCTGCTACTCACTCGTGTCTACT       730
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   728  GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAATGATAAAAGATTTTACCTGAAATG     787

Db   731  GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAATGATAAAAGATTTTACCTGAAATG     790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   788  ATTTTGGCTTCATGATCTTTCACTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC    847

Db   791  ATTTTGGCTTCATGATCTTTCACTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC    850
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   848  GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAAACTAGTACTCTATAGTG        907

Db   851  GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAAACTAGTACTCTATAGTG        910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   908  GCCTGGAAGCCCTTTGGGGAGTTGGAGACACCTACTTTCGATACGAGATGTGGAGA        967

Db   911  GCCTGGAAGCCCTTTGGGGAGTTGGAGACACCTACTTTCGATACGAGATGTGGAGA        970
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   968  AGGCATTTATGAAGGCTAGCTCTGAAACTTCAGTAGAGAACACCTACTTTCTTTACTTG      1027

Db   971  AGGCATTTATGAAGGCTAGCTCTGAAACTTCAGTAGAGAACACCTACTTTCTTTACTTG      1030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1028  TATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATAGGTTACACATCTTCATCTGTTC    1087

Db  1031  TATCAAATCAAAATGGAAATATGTACACATCTTCAGTATATAGGTTACACATCTTCATCTGTTC    1090
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1088  TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTTTCTTATGGTT      1147

Db  1091  TAGCACAGTACTCACTCAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTTTCTTATGGTT      1150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1148  CTGGTTTGGCTGTGCCACTCGTACTCTCTTAAAGTCACACAAGATGCTACCGGGGTCTG      1207

Db  1151  CTGGTTTGGCTGTGCCACTCGTACTCTCTTAAAGTCACACAAGATGCTACCGGGGTCTG      1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1208  CTCTTGATAAAATAACAGACAAGTTATGTGATCTTAAAATCAAAGGCTTGGAGAACTG      1267

Db  1211  CTCTTGATAAAATAACAGACAAGTTATGTGATCTTAAAATCAAAGGCTTGGAGAACTG      1270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1268  GTGTGGCACCAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGGACGAGGACAGCACCCAT     1327

Db  1271  GTGTGGCACCAGATGTCTTGGCTGAAAAACATGAAGCTCAGAGGACGAGGACAGCACCCAT     1330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1328  TCAACTATATTCCCCAGGGGTTCAATAGATTCACTCTTCTTTGAAGGAAGGTGGTACTTAGTTA    1387

Db  1331  TCAACTATATTCCCCAGGGGTTCAATAGATTCACTCTTCTTTGAAGGAAGGTGGTACTTAGTTA    1390
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1388  GGGTGGATGAAAAGCACAGAAGAACTTAGCCTCGGGGTCCCACTCCAAAATGATGACTT      1447

Db  1391  GGGTGGATGAAAAGCACAGAAGAACTTAGCCTCGGGGTCCCACTCCAAAATGATGACTT      1450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1448  TGGATGAAGGAGTGAGGAGGACTTGGCATTCAAACAATAGCAACTGAACTGAACTGAGC     1507

Db  1451  TGGATGAAGGAGTGAGGAGGACTTGGCATTCAAACAATAGCAACTGAACTGAACTGAGC     1510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1508  CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCAGAACCTGAAGCAGCTGTCATTAGTA    1567

Db  1511  CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCAGAACCTGAAGCAGCTGTCATTAGTA    1570
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1568  ATGGGGAACATTAAGATACTCTGTGGAGGTGCCAAGACTTCAGGGTGGGGTGGGCATGGGGT     1627

Db  1571  ATGGGGAACATTAAGATACTCTGTGGAGGTGCCAAGACTTCAGGGTGGGGTGGGCATGGGGT     1630
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1628  GGGGGTATGGGACAGTTGG      1647

Db  1631  GGGGGTATGGGACAGTTGG      1650
```

```
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-205

Query Match   68.4%;    Score 1370;   DB 13;   Length 1650;
Best Local Similarity   92.0%;   Pred. No. 0;
Matches 1509;   Conservative   0;   Mismatches   5;   Indels 126;   Gaps 1;
```

```
Qy      134  GCTCTTTCACCAATGCCTGGATCACTCTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG   193
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       11  GCTCTTTCACCAATGCCTGGATCACTCTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG    70

Qy      194  TGGGAAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGG   253
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       71  TTGGGATTGTTGCCCTTGAGATCTATTTTCCTTCAATATGTTGATCAAGCAGAGTTGG     130

Qy      254  AAAAATATGATGGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG   313
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      131  AAAAATATGATGGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG   190

Qy      314  GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGATGATGTGTTTCAGAATCTTA   373
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      191  GCTTCTGCACAGATAGAGAAGAGATATTAACTCTCTTTGATGATGTGTTTCAGAATCTTA   250

Qy      374  TGGAGAGAAATAACCTTTCCTTATCTATGATTTCCATTGGGCGGCTGGAAGTTGGAACAA   433
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      251  TGGAGAGAAATAACCTTTCCTTATCTATGATTTCCATTGGGCGGCTGGAAGTTGGAACAGAGACAA   310

Qy      434  TCATCGACAAATCAAAGTTCTGTGAAGACTAATTGATGATGTCAGCTGTTGAAGAGTCTGGGA   493
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      311  TCATCGACAAATCAAAGTTCTGTGAAGACTAATTGATGATGTCAGCTGTTGAAGAGTCTGGGA   370

Qy      494  ATACAGATATAGAGAGGAATCGACACAACTAATGCATGCTGTATGAGGCACAGAGTGCTGCT   553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      371  ATACAGATATAGAGAGGAATCGACACAACTAATGCATGCTATGAGGCACAGCTGCTGTGCT   430

Qy      554  TCAATGCTGTTAACTGGATTGAGTGAGTCCAGCTGTCTTGGGAT--------           591
             |||||||||||||||||||||||||||||||||||||||||||
Db      431  TCAATGCTGTTAACTGGATTGAGTGAGTCCAGCTCTTTGGGATGGACGGTATGCCCTGTAGTTG   490

Qy      592  ------------------------------------------GGGCTTCGTGGGACAC   607
                                                       ||||||||||||||||
Db      491  CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGTTGGACGCAGGGGCTTCGTGGGACAC   550

Qy      592  TAGCTCTCGCTAATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACAC   667
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      551  GGGGGGTTATGGGAAACAGTTGG   610

Qy      608  ATATGCAACATGCCTATGATATGCTAGCTACATCTGAAATATCCTATAGTTAG   667
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      611  ATATGCAACATGCCTATGATTTTTACAGCCTGATATGCTATCTCTGAATATCCTATAGTTAG   670

Qy      668  ATGGAAAACTCCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT   727
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      671  ATGGAAAACTCCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACT   730

Qy      728  GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAGAGATTTACTTGAATG   787
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      731  GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATAAGAGATTTACTTGAATG   790

Qy      788  ATTTTTGGCTTCATGATCTTTCACTCACCATATATGTTGTAAAACTGGTTCAGAAATCTCTAGCTC   847
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      791  ATTTTTGGCTTCATGATCTTTCACTCACCATATGTTGTAAAACTGGTTCAGAAATCTCTAGCTC   850

Qy      848  GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAAGAATAAAAAATAGTATCTATAGTG   907
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      851  GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAAGAATAAAAAATAGTATCTATAGTG   910

Qy      908  GCCTGGGAAGCCTTTGGGGGATTGTAAAATTAGAGAAGACACCTACTTTGATAGAGAGATGTGGAGA   967
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      911  GCCTGGGAAGCCTTTGGGGGATTGTAAAATTAGAGAAGACACCTACTTTGATAGAGAGATGTGGAGA   970

Qy      968  AGGCATTTATGAAGGCTAGCTTCTTCAGTCGTCAGAATACAAAGGCAATCTTTACTTG   1027
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      971  AGGCATTTATGAAGGCTAGCTTCTTCAGTCGTCAGAATACAAAGGCAATCTTTACTTG   1030

Qy     1028  TATCAAAATCAAAAATGGAAAATATGTACACATCTTCAGTGTATATGGTTCCCTTGCATCTGTTC   1087
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1031  TATCAAAATCAAAAATGGAAAATATGTACACATCTTCAGTGTATATGGTTCCCTTGCATCTGTTC   1090

Qy     1088  TAGCACAGTACTCACCTCAGCAATTAGCACGGGAAGGAGAAGAAAATTGGAGTGTTTTTCTTATGGTT   1147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1091  TAGCACAGTACTCACCTCAGCAATTAGCACGGGAAGGAGAAGAAAATTGGAGTGTTTTTCTTATGGTT   1150

Qy     1148  CTGGTTTGGCTGCTGCCACTCTGTACTCTCTTAAAGTCACACACAAGAGATCTTACACCGGGGTCTG   1207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1151  CTGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAAGAGATGGAGTGGGGTCTG   1210

Qy     1208  CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAAATCAAGGCTTCAAGAACTG   1267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1211  CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAAATCAAGGCTTCAAGAACTG   1270

Qy     1268  GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAAGCTCAGAGGAGAGGACACCCATCATTGG   1327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1271  GTGTGGCACCAGATGTCTTCGCTGAAAACATGAAAGCTCAGAGGAGAGGACACCCATCATTGG   1330

Qy     1328  TCAACTATATTCCCCAGGGGTTCAAATAGAATTCAAGTTCACTCTTTGAAGGAGGAACGTGGTGGTTACTTAGTTA   1387
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1331  TCAACTATATTCCCCAGGGGTTCAAATAGAATTCAAGTTCACTCTTTGAAGGAGGAACGTGGTGGTTACTTAGTTA   1390

Qy     1388  GGGTGGATGAGAAAAGCACACAGAAGAACTTACCTCGCGGTCCCAAATGCTCCAAATCTCCAAATGAGACACTT   1447
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1391  GGGTGGATGAGAAAAGCACACAGAAGAACTTACCTCGCGGTCCCCACTCCAAATGAGACACTT   1450

Qy     1448  TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCATCAATGACTGCCAAGCCCTG   1507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1451  TGGATGAAGGAGTAGGACTTGTGCATTCAAACATAGCATCAATGACTGCCAAGCCCTG   1510

Qy     1508  CCAAGAAAAGTACCAAGACTCCCTGCCACACAGCAGAAACCTGAACGAGCCTGTCATTAGTA   1567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1511  CCAAGAAAAGTACCAAGACTCCCTGCCACACAGCAGAAACCTGAACGAGCTCATTAGTA   1570

Qy     1568  ATGGGGGAACATTAAGATACTTGTGAGGGTGGCAAGACTTCAGGGGTGGGGGTGGGGGCATGGGGT   1627
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1571  ATGGGGGAACATTAAGATACTTGTGAGGGTGGCAAGACTTCAGGGGTGGGGGTGGGGGCATGGGGT   1630

Qy     1628  GGGGGTTATGGGGAACAGTTGG   1647
             ||||||||||||||||||||||
Db     1631  GGGGGTTATGGGAAACAGTTGG   1650
```

```
RESULT 12
US-10-307-817-359
; Sequence 359, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 359
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1581)
US-10-307-817-359
```

```
Query Match             68.4%;  Score 1370;  DB 13;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 0;  Mismatches 0;
Matches 1509; Conservative    0; Mismatches    5; Indels 126; Gaps    1;

Qy    134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70

Qy    194 TGGGAATTGTTGCCCTGAGATCTATTTCCTTCTTCAAATAGTTGTGATCAAGCAGAGTTGG 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TTGGGAATTGTTGCCCTGAGATCTATTTCCTTCTTCAAATAGTTGTGATCAAGCAGAGTTGG 130

Qy    254 AAAAATATGATGGTGTAGAGTGCTGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 AAAAATATGATGGTGTAGAGTGCTGGAAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGG 190

Qy    314 GCTTCTGCACAGATAGAGGAAGAGAATATTAACTCTCTTTGCATGACTGTGGGAACAGACA 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 GCTTCTGCACAGATAGAGGAAGAGAATATTAACTCTCTTTGCATGACTGTGGGAACAGACA 250

Qy    374 TGGAGAGAAATAAACCTTTCCTATGAATTGCATTGGGCGGCCGGCGGAAGTTGGAACAGACA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 TGGAGAGAAATAAACCTTTCCTATGAATTGCATTGGGCGGCCGGCGGAAGTTGGAACAGACA 310

Qy    434 TCATCGACAAAATCAAAAGTCTGTGTGAAGACTAAATTGATCAGCGTGTTTGAGAGTCTGGA 493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCATCGACAAAATCAAAAGTCTGTGTGAAGACTAAATTGATCAGCGTGTTTGAGAGTCTGGA 370

Qy    494 ATACAGATATAGGAAGGAATCGACACACAACTAATGCATGGAGGCACAGCTGCTGCTGCT 553
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 ATACAGATATAGGAAGGAATCGACACACAACTAATGCATGGAGGCACAGCTGCTGCTGCT 430

Qy    554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------------ 591
          ||||||||||||||||||||||||||||
Db    431 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTG 490

Qy    592 ------------------------------GGGCTTCGTTGGGGACAC 607
                                        ||||||||||||||||
Db    491 CAGGAGGATAATTGCTGTATATGCCACAGGAAATGCTAGACCTACAGGGTGGAGTTGGGACGAG 550

Qy    592 --------------------------------GGGCTTCGTTGGGGACAC 607
                                          ||||||||||||||||
Db    551 TAGCTCTGCTAATTGGGCCAAAATGCTCCTTTAATTTTTGAACAGAGACC 610

Qy    608 ATATGCAACATGCCTATGAGTTTTTTACAAGCCTGATTATTGCTATCTGAAATATCCTATAGTTAG 667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACATGCCTATGAGTTTTTTACAAGCCTGATTATTGCTATCTGAAATATCCTATAGTTAG 670

Qy    668 ATGGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAGCCGTCACTCGTGTCTACT 727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGAGCCGTCACTCGTGTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAATGATTGATAAAGATTTTACCTTGAATG 787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAATGATTGATAAAGATTTTACCTTGAATG 790

Qy    788 ATTTTCGGCTTCATGATCTTCACTCACCATATTGTAAAACTGGTTCAGAAAATCTTCTAGCTC 847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTCGGCTTCATGATCTTCACTCACCATATTGTAAAACTGGTTCAGAAAATCTTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATCAGTAGAAAAGATACATGTATTCTATAGTG 907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATACAGAGATAAAAATAGTATTCTATAGTG 910

Qy    908 GCCTGGAAGGCCTTTGGGGATGTTAAATTATTAGAGACAACACCTACTTTGATAGAGATGTGGAGA 967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGAAGGCCTTTGGGGATGTTAAATTATTAGAGACAACACCTACTTTGATAGAGATGTGGAGA 970

Qy    968 AGGCATTTATGAAGGGTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTTACTTG 1027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAAGGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTTACTTG 1030

Qy   1028 TATCAAATCAAAATGGAAAATGTACACATCTTCAGTATATGTATGGTTCCCTGGCATCGTTC 1087
```

```
                                                        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAATCAAAATGGAAAATGTACACATCTTCAGTATATGTATGGTTCCCTGGCATCGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCACAGGGACAGAGAATTGGAGTGTTTTCTTATGGTT 1147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCACAGGGACAGAGAATTGGAGTGTTTTCTTATGGTT 1150

Qy   1148 CTGGGTTTGGCTGCCACTCGTGACTCCTCTTTAAAGTCACACAAAGATGTCACCACGGGTCTG 1207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGGTTTGGCTGCCACTCGTGACTCCTCTTTAAAGTCACACAAAGATGTCACCACGGGTCTG 1210

Qy   1208 CTCTTGATAAAAATAAACACGCAAGTTTAATGTGATCTTAAATCAAGGCTTGAAATCAAGAACTG 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAAATAAACACGCAAGTTTAATGTGATCTTAAATCAAGGCTTGAAATCAAGAACTG 1270

Qy   1268 GTGTGGCACCAGATGTCTTCGTGAAAACATGAAAGCTCAGAGAGGACACCCATCATTGG 1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCAGATGTCTTCGTGAAAACATGAAAGCTCAGAGAGGACACCCATCATTGG 1330

Qy   1328 TCAACTATATTCCCCAGGGTTCAAATAGATTCACTCCTTTGAAGGAAGGTGGTACTTAGTTA 1387
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAAGGTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAGCCACAGAAGAAGAACTTACGCTGCGGGTGCCCACTCCAAATGATGACACTT 1447
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGATGAAAAGCCACAGAAGAAGAACTTACGCTGCGGGTGCCCACTCCAAATGATGACACTT 1450

Qy   1448 TGGATGAAGGAGTGGGACTTGTTGCGATTCAAACATTAGCACATCAAGCATATTCAAAGCCCTG 1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTGGGACTTGTTGCGATTCAAACATTAGCACATCAAGCATATTCAAAGCCCTG 1510

Qy   1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCCAGTCAGTCTGCATTAGTA 1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCCAGTCAGTCTGCATTAGTA 1570

Qy   1568 ATGGGGAAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGCGGGGTGGGGGT 1627
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGCGGGGTGGGGGT 1630

Qy   1628 GGGGGGTTATGGGAACAGTTGG 1647
          ||||||||||||||||||||||
Db   1631 GGGGGGTTATGGGAACAGTTGG 1650
```

```
RESULT 13
US-10-307-817-371
; Sequence 371, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-502C
;  CURRENT APPLICATION NUMBER: US/10/307,817
;  CURRENT FILING DATE: 2002-12-02
;  NUMBER OF SEQ ID NOS: 682
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 371
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (22)..(1581)
US-10-307-817-371
```

```
Query Match             68.4%;  Score 1370;  DB 13;  Length 1650;
Best Local Similarity   92.0%;  Pred. No. 0;  Mismatches 0;
Matches 1509; Conservative    0; Mismatches    5; Indels 126; Gaps    1;

Qy    134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATG 70
```

```
RESULT 14
US-10-307-817-445
; Sequence 445, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 445
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-307-817-445

  Query Match             68.4%;  Score 1370;  DB 13;  Length 1650;
  Best Local Similarity   92.0%;
  Matches 1509;  Conservative   0;  Mismatches   5;  Indels 126;  Gaps   1;
```

```
Qy    134 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAGAGCTTGCTGGCCAAAAGATG 193
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db     11 GCTCTTTCACCATGCCTGGATCACTTCCTTTGAATGCAGAGAGCTTGCTGGCCAAAAGATG 70

Qy    194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATAATATGTTGATCAAGCAGAGTTGG 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAATAATATGTTGATCAAGCAGAGTTGG 130
```

(Sequence alignment data — query (Qy) vs database (Db) nucleotide sequences)

```
Qy    374 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGGGGGAGTTGGAACAGCAGAGACAA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 TGGAGAGAAATAACCTTTCCTATGATTGCATTGGGCGGCGGGGGGAGTTGGAACAGCAGAGACAA 310

Qy    434 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTGATTGCATGCACCTGTTTGAAGAGTCTGGGA 493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 TCATCGACAAATCAAAGTCTGTGAAGACTAAATTGATTGCACGCACCTGTTTGAAGAGTCTGGGA 370

Qy    494 ATACAGATATAGAAGGAATCGACACAACTAAATGCATCGCTATGGAGGAGCCAGCAGCTGCTGTCT 553
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 ATACAGATATAGAAGGAATCGACACAACTAAATGCATCGCTATGGAGGAGCCAGCAGCTGCTGTCT 430

Qy    554 TCAATGCTGTTAACTGGATTGAGTCCAGTCCAGTCACAC------------------------ 591
          ||||||||||||||||||||||||||||||||||||||||
Db    431 TCAATGCTGTTAACTGGATTGAGTCCAGTCCAGTCACAC 490
                                                                         591

Qy    592 ------------------------------------------------------------- 591
Db    491 CAGGAGGATATTGCTGTGTATATGCCACACAGGGAAATGCTAGACCTACAGGTTGGGAGTGGAGCAG 550

Qy    592 ---------GGGCTTCGTTCGTGGGGACAC 607
                   |||||||||||||||||||||
Db    551 TAGCTCTGCTAATTGGGGCCAAAATGCTCCTTTAATTTTTGAACGAGAGGGCTTCGTTCGTGGGGACAC 610

Qy    608 ATATGCAACACTGCCTATGATTTTTACCTATCCTGATATGCTATCCTATAGTATATCCTATAGTAG 667
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCCTATTCTGAATATCCTATAGTAG 670

Qy    668 ATGGAAAACTCTCCATACAGTGCTACTCCAGTGCATTAGACCGGCTTGCTCCTACTCTGTTCTACT 727
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    671 ATGGGAAACTCTCCATACAGTGCTACTCCAGTGCATTAGACCGCTGCTTCTGCTACTCTGTTCTACT 730

Qy    728 GCAAAAAGATCCATGCCCAGTGGCAGAAAAGAGGGAAATGATAAAAGATTTTTACCTTGAAATG 787
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 GCAAAAAGATCCATGCCCAGTGGCAGAAAAGAGGGAAATGATAAAAGATTTTTACCTTGAAATG 790

Qy    788 ATTTTGGCTTCATGATCTTTCACTCACCATATTGTAACTGGTTCAGAAATCTCTAGCTC 847
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 ATTTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 850

Qy    848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATGAGAATAAAAAGTAGTATCTATAGTG 907
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAAAAAGTAGTATCTATAGTG 910

Qy    908 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGGACACCTTACTTTGATAGAGATGTGGAGA 967
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCCTGGAAGCCTTTGGGGATGTTAAATTAGGAAGACACCTTACTTTGATAGAGATGTGGAGA 970

Qy    968 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGTCAGAAAACAAAGGCATCTTTACTTG 1027
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 AGGCATTTATGAAGGCTAGCTCTGAACTCTTCAGTGTCAGAAAACAAAGGCATCTTTACTTG 1030

Qy   1028 TATCAAATCAAAATGGAAAATAGTACACATCTTCAGTGTATGGTTCCCTTGCATCTGTTC 1087
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 TATCAAAATCAAAATGTACACATCTTCAGTGTATGGTTCCCTTGCATCTGTTC 1090

Qy   1088 TAGCACAGTACTCACCTCAGCAATTAGCACGGAGAGAGAAGAATTGGAGTGTTTTCTTATGGTT 1147
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 TAGCACAGTACTCACCTCAGCAATTAGCACGGAGAGAGAAGAATTGGAGTGTTTTCTTATGGTT 1150

Qy   1148 CTGGTTTGGGCTGCCACTCACTGCTGTACTCCTGTACTCTCTTAAAGTCACACAAGATCACACCGGGGTCTG 1207
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 CTGGTTTGGGCTGCCACTCACTGCTGTACTCCTGTACTCTCTTAAAGTCACACAAGATCACACCGGGGTCTG 1210

Qy   1208 CTCTTGATAAATACACGAAGTTTATGTTGATCTTAAATCAACAAGAGACTG 1267
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 CTCTTGATAAAATAACAGCAAGTTTATGTTGATCTTAAATCAACAAGAGACTG 1270

Qy   1268 GTGTGGCACCAGATGTCTTCGCCTGAAAAACATGAAGCTCAGAGAGAGAGCACCCATCATTTGG 1327
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 GTGTGGCACCAGATGTCTTCGCCTGAAAAACATGAAGCTCAGAGAGAGAGCACCCATCATTTGG 1330
```

RESULT 15
US-10-354-358-55
; Sequence 55, Application US/10354358
; Publication No. US20030157082A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc
; APPLICANT: Hunter, John Joseph
; APPLICANT: Macbeth, Kyle J.
; APPLICANT: Tsai, Fong-Ying
; APPLICANT: Lesson, Andrea
; APPLICANT: Lightcap, Eric S.
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: CANCER USING 140, 1470, 1686, 2089, 2427, 3702, 5891, 6428,
; TITLE OF INVENTION: 7181, 7660, 25641, 69583, 49863, 8897, 1682, 17667, 9235,
; TITLE OF INVENTION: 3703, 14171, 10359, 1660, 1450, 18894, 2088, 32427, 2160,
; TITLE OF INVENTION: 9252, 9389, 1642, 85269, 10297, 1584, 9525, 14124, 4469,
; TITLE OF INVENTION: 8990, 2100, 9288, 64698, 10480, 20893, 33230, 1586, 9943,
; TITLE OF INVENTION: 16334, 68862, 9011, 14031, 6178, 21225, 1420, 32236, 2099,
; TITLE OF INVENTION: 2150, 26583, 2784, 8941, 9811, 27444, 50566 OR 66428 MOLECULES
; FILE REFERENCE: MPI02-020P1RNOMNIM
; CURRENT APPLICATION NUMBER: US/10/354,358
; CURRENT FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: US/10/354,358
; PRIOR FILING DATE: 2003-01-30
; PRIOR APPLICATION NUMBER: US 60/353,600
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: US 60/364,517
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/371,075
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 60/371,507
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: US 60/372,984
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: US 60/374,194
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/382,995
; PRIOR FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/385,023
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/388,853
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/389,395
; PRIOR FILING DATE: 2002-06-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 55

```
Qy   1328 TCAACTATATTCCCCAGGGTTCAATAGAATTCACTCTTTGAAGGAAGGAACGTTGGTACTTAGTTA 1387
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1331 TCAACTATATTCCCCAGGGTTCAATAGAATTCACTCTTTGAAGGAAGGAACGTTGGTACTTAGTTA 1390

Qy   1388 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGGCTCCCACTCCCAAAATGATCACACTT 1450
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 GGGTGGATGAAAAGCACAGAGAGAACTTACGCTCGGCGGCTCCCACTCCCAAAATGATCACACTT 1450

Qy   1448 TGGATGAAGGAGTAGGAGACTTTGTGCATTCAAACATAGCAACTGAGCAATATTCCAAGCCCTG 1507
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 TGGATGAAGGAGTAGGAGACTTTGTGCATTCAAACATAGCAACTGAGCAATATTCCAAGCCCTG 1510

Qy   1508 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCCAGGAAGCAGCAGCTGTTGTCATTAGTA 1567
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1511 CCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCCAGGAAGCAGCAGCTGTTGTCATTAGTA 1570

Qy   1568 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGTGGGGCATGGGGT 1627
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1571 ATGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGTGGGGCATGGGGT 1630

Qy   1628 GGGGGTATGGGAACAGTTGG 1647
          ||||||||||||||||||||
Db   1631 GGGGGTATGGGAACAGTTGG 1650
```

```
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (22)...(1584)
US-10-354-358-55

Query Match   68.4%;  Score 1370;  DB 15;  Length 1650;
Best Local Similarity  92.0%;  Pred. No. 0;
Matches 1509;  Conservative  0;  Mismatches 0;  Indels 126;  Gaps 1;
```

```
Qy   134 GCTTCTTCACCATGCCTGGATCACTCACTTCCTTTGAATGCAGAAGCTTGCTGGGCCAAAAGATG 193
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    11 GCTTCTTCACCATGCCTGGATCACTCACTTCCTTTGAATGCAGAAGCTTGCTGGGCCAAAAGATG 70

Qy   194 TGGGAAATTGTTGCCCTTGAGATCTATTTTTCCTTCTCAATATAGTTGATCAAGCAAGAGTTGG 253
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    71 TTGGGATTGTTGCCCTTGAGATCTATTTTTCCTTCTCAATATATGTTGATCAAGCAGAGTTGG 130

Qy   254 AAAAATATGATGGTGTAGATGCTGGAAAGTATAACCATTGGCCTTGGGCCAGGCCAAGATGG 313
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   131 AAAAATATGATGGTGTAGATGCTGGAAAGTATAACCATTGGCCTTGGGCCAGGCCAAGATGG 190

Qy   314 GCTTCTGCACAGATAGAGAGAAGAGATATTAACTCTCTCTTTGCATGATGTGTTCAGAATCTTA 373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   191 GCTTCTGCACAGATAGAGAGAAGAGATATTAACTCTCTCTTTGCATGATGTGTTCAGAATCTTA 250

Qy   374 TGGAGGAGAAATAACCTTTCCTTTCCTTATGATTCCATTCGGGCGGCGGCCTGGAAGTTGGAACAGAGACAA 433
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   251 TGGAGGAGAAATAACCTTTCCTATGATTCCATTGGGCGGCGGCCTGGAAGTTGGAACAGAGACAA 310

Qy   434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCCAGCTGTTGAAGAGTTCTGGGA 493
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCCAGCTGTTTGAAGAGTTCTGGGA 370

Qy   494 ATACAGATATAGAAGGAATCGACACACAACTAATGCATGCTTATATGAGGAGGCACAGCTGCTGTCT 553
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 ATACAGATATAGAAGGAATCGACACACAACTAATGCATGCTTATATGGAGGAGGCACAGCTGCTGTCT 430

Qy   554 TCAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------ 591
         |||||||||||||||||||||||||||||||||
Db   431 TCAAATGCTGTTAACTGGATTGAGTCCAGCTCTTGGGAT--- 490

Qy   592 ------------------------------------------------ 591
Db   491 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCCTAGACCTACAGGTGGGAGTTGGGAGCAG 550

Qy   592 ---------------GGGCTTCGTGGGGACAC 607
                        ||||||||||||||||
Db   551 TAGCCTCTGCTAATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGGACAC 610

Qy   608 ATATGCAACATGGCCTATGATTTTTACAAGCCTGATATGCCTATCTGAAATATCCTATAGTAG 667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   611 AATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCCTATCTGAAATATCCTATAGTAG 670

Qy   668 ATGGAAAACTCTCCATACACAGTGCTTACCTCCAGTGCATTACAGACCTGCTACTCTCGTCTACT 727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   671 ATGGAAAACTCTTCCATACACAGTGCTTACCTCAGTGCATTAGACCCGCTGCTCTTCGTCTACT 730

Qy   728 GCAAAAAGATCCATGCCCAGTGCAGAAATGATCAGAGGGAAAATCAATAAAAGATTTTACCTTGAATG 787
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731 GCAAAAAGATCCATGCCCAGTGCAGAAATGATCAGAGGGAAAATCAATAAAAGATTTTACCTTGAATG 790

Qy   788 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTTGTTAAACTGGTTGCAGAAATCTTAGCTC 847
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791 ATTTTGGCTTCATGATCTTTCACTCACCCATATTGTTGTAAACTGGTTCAGAAATCTCTAGCTC 850

Qy   848 GGATGTTGCTGAATGACTTACTTAATGACCAGAATAGAGAGATAAAAAATAGTATCTATAGTG 907
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGAGATAAAAAATAGTATCTATAGTG 910

Qy   908 GCCTGGGAAGCCTTTGGGGATGTTGTAAAATTAGAGAGACACCTACTTTGATAGAGATGTGGAGA 967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   911 GCCTGGGAAGCCTTTGGGGATGTTGTAAAATTAGAGAGACACCTACTTTGATAGAGATGTGGAGA 970

Qy   968 AGGGCATTTATGAAGGCTAGCTTCAGTCAGAAATCAAAAGGCATCTTTACTTG 1027
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   971 AGGGCATTTATGAAGGCTAGCTTCAGTCAGAAATCAAAAGGCATCTTTACTTG 1030

Qy  1028 TATCAAAATCAAAAATGGAAATAGTGTACACATCTTCAGTGTATATGGTTCCCTTGCATCTGTTC 1087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1031 TATCAAAATCAAAAATGGAAATAGTGTACACATCTTCAGTGTATATGGTTCCCTTGCATCTGTTC 1090

Qy  1088 TAGCACAGTACTCACCTCAGCCAATTAGCAGGGAAGGAAGGAAATTGGAAGTGGAAGTGTTTTTTCTTTATGGTT 1147
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1091 TAGCACAGTACTCACCTCAGCCAATTAGCAGGGAAGGAAGGAAATTGGAGTGTTTTTCTTTATGGTT 1150

Qy  1148 CTGGTTTGGCTGCGCCACTCGTACTCTCTTAAAGTCACACAAGAGTCTCGG 1207
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1151 CTGGTTTGGCTGCGCCACTCGTACTCTCTTAAAGTCACACAAGAGATCGCTG 1210

Qy  1208 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAAACTG 1267
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211 CTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAAACTG 1270

Qy  1268 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGGAGGACACCCATCATTGG 1327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1271 GTGTGGCACCAGATGTCTTCGCTGAAAAACATGAAGCTCAGAGGAGGACACCCATCATTGG 1330

Qy  1328 TCAACTATATTCCCCAGGGTTCAATAGAATCACTCTTTGAAGGAACGTGGTACTTTAGTTA 1387
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1331 TCAACTATATTCCCCAGGGTTCAATAGAATCACTCTTTGAAGGAACGTGGTACTTTAGTTA 1390

Qy  1388 GGGGTGGATGAAAAGCACAGAAGAACTTACGGCTCGGGCTCCACTCCAAATGATGACACTT 1447
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1391 GGGTGGATGAAAAGCACAGAAGAACTTACGGCTCGGGCTCCACTCCAAATGATGACACTT 1450

Qy  1448 TGGATGAAGGAGTAGGGACTTGTGCAATTCAAAACATAGCTGAGCATATTCCAAGCCCTG 1507
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1451 TGGATGAAGGAGTAGGGACTTGTGCAATTCAAAACATAGCTGAGCATATTCCAAGCCCTG 1510

Qy  1508 CCAGAAACTTACCAAGACTCCCCTGCCACAGCAGCAGCCTGAGCAGCTGTCATTGTTCATTAGTA 1567
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1511 CCAGAAAGTTACCAAGACTCCCCTGCCACAGCAGCAGCCTGAGCAGCTGTCATTATTAGTA 1570

Qy  1568 ATGGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGTGGGGT 1627
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1571 ATGGGGGAACATTAAGATACTTCTGTGAGGTGCAAGACTTCAGGGTGGGGTGGGGCATGGGGGT 1630

Qy  1628 GGGGGTATGGGGAACAGTTGG 1647
         ||||||||||||||||||
Db  1631 GGGGGTATGGGGAACAGTTGG 1650
```

```
Search completed: June 24, 2004, 13:16:16
Job time : 860 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 13:20:33 ; Search time 883 Seconds
152.827 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV...........PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1163542 seqs, 282313646 residues

(without alignments)
Total number of hits satisfying chosen parameters: 1163542

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :
Published_Applications_AA:*
1: /cgn2_6/ptcdata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptcdata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptcdata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptcdata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptcdata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptcdata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptcdata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptcdata/2/pubpaa/US09A_PUBCOMB.pep:*
9: /cgn2_6/ptcdata/2/pubpaa/US09B_PUBCOMB.pep:*
10: /cgn2_6/ptcdata/2/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptcdata/2/pubpaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptcdata/2/pubpaa/US10A_PUBCOMB.pep:*
13: /cgn2_6/ptcdata/2/pubpaa/US10B_PUBCOMB.pep:*
14: /cgn2_6/ptcdata/2/pubpaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptcdata/2/pubpaa/US10_NEW_PUB.pep:*
16: /cgn2_6/ptcdata/2/pubpaa/US60_NEW_PUB.pep:*
17: /cgn2_6/ptcdata/2/pubpaa/US60_PUBCOMB.pep:*
18: /cgn2_6/ptcdata/2/pubpaa/US60_PUBCOMB.pep:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2511	100.0	478	13	US-10-193-295-2	Sequence 2, Appl
2	2511	100.0	478	15	US-10-622-516-2	Sequence 2, Appli
3	2480	98.8	520	12	US-10-236-417-182	Sequence 182, App
4	2480	98.8	520	12	US-10-236-417-184	Sequence 184, App
5	2480	98.8	520	12	US-10-236-417-186	Sequence 186, App
6	2480	98.8	520	12	US-10-236-417-190	Sequence 190, App
7	2480	98.8	520	12	US-10-236-417-194	Sequence 194, App
8	2480	98.8	520	12	US-10-236-417-198	Sequence 198, App
9	2480	98.8	520	12	US-10-236-417-202	Sequence 202, App
10	2480	98.8	520	12	US-10-236-417-204	Sequence 204, App
11	2480	98.8	520	12	US-10-236-417-206	Sequence 206, App
12	2480	98.8	520	12	US-10-236-417-208	Sequence 208, App
13	2480	98.8	520	12	US-10-307-817-360	Sequence 360, App
14	2480	98.8	520	12	US-10-307-817-370	Sequence 370, App
15	2480	98.8	520	12	US-10-307-817-372	Sequence 372, App
16	2480	98.8	520	12	US-10-307-817-374	Sequence 374, App
17	2480	98.8	520	12	US-10-307-817-467	Sequence 467, App
18	2480	98.8	520	12	US-10-307-817-468	Sequence 468, App
19	2480	98.8	520	12	US-10-041-018-292	Sequence 292, App
20	2480	98.8	520	13	US-10-193-295-4	Sequence 4, Appli
21	2480	98.8	520	14	US-10-354-358-56	Sequence 56, Appl
22	2480	98.8	520	15	US-10-622-516-4	Sequence 4, Appli
23	2480	98.8	529	12	US-10-307-817-192	Sequence 192, App
24	2480	98.8	529	12	US-10-307-817-364	Sequence 364, App
25	2480	98.8	519	12	US-10-307-817-188	Sequence 188, App
26	2475	98.6	519	12	US-10-307-817-362	Sequence 362, App
27	2475	98.6	524	12	US-10-236-417-196	Sequence 196, App
28	2475	98.6	524	12	US-10-307-817-366	Sequence 366, App
29	2475	98.6	524	12	US-10-236-417-212	Sequence 212, App
30	2475	98.6	524	12	US-10-307-817-376	Sequence 376, App
31	2475	98.6	525	12	US-10-236-417-200	Sequence 200, App
32	2475	98.6	525	12	US-10-236-417-210	Sequence 210, App
33	2475	98.6	525	12	US-10-307-817-358	Sequence 358, App
34	2475	98.6	525	12	US-10-307-817-368	Sequence 368, App
35	2439	97.1	518	13	US-10-193-295-5	Sequence 5, Appli
36	2439	97.1	518	15	US-10-622-516-5	Sequence 5, Appli
37	2362	94.1	520	12	US-10-205-331-14	Sequence 14, Appl
38	2362	94.1	520	12	US-10-307-817-364	Sequence 313, App
39	2352	93.7	522	12	US-10-041-018-314	Sequence 314, App
40	2097	83.5	520	12	US-10-041-018-315	Sequence 315, App
41	1513	60.3	508	12	US-10-041-018-296	Sequence 296, App
42	1513	60.3	508	14	US-10-205-823-174	Sequence 174, App
43	1513	60.3	508	14	US-10-177-293-208	Sequence 208, App
44	1494	59.5	508	12	US-10-041-018-316	Sequence 316, App
45	1493	59.5	480	12	US-10-041-018-284	Sequence 284, App

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-10-193-295-2
Sequence 2, Application US/10193295
Publication No. US20020173018A1
GENERAL INFORMATION:
APPLICANT: GONG, Pangcheng et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
TITLE OF INVENTION: THEREOF
FILE REFERENCE: CL0011195DIV
CURRENT APPLICATION NUMBER: US/10/193,295
CURRENT FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: 08/819,993
PRIOR FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 478
TYPE: PRT
ORGANISM: Human
US-10-193-295-2

Query Match 100.0%;
Best Local Similarity 100.0%;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Score 2511; DB 13; Length 478;
Pred. No. 5.2e-241;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQAXMGFCT 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQAXMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180

```
Qy  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240

Qy  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300

Qy  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSSALDKIT  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSSALDKIT  360

Qy  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420

Qy  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
```

RESULT 2
US-10-622-516-2
Sequence 2, Application US/10622516
; Publication No. US20040018545A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00119SDIV2
; CURRENT APPLICATION NUMBER: US/10/622,516
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
; SOFTWARE: FastSEQ for Windows Version 4.0

Query Match 100.0%; Score 2511; DB 15; Length 478;
Best Local Similarity 100.0%; Pred. No. 5.2e-241;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-10-622-516-2

```
Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180

Qy  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240

Qy  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300

Qy  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSSALDKIT  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSSALDKIT  360

Qy  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420

Qy  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
```

RESULT 3
US-10-236-417-182
Sequence 182, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 182
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-236-417-182

Query Match 98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

```
Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------  149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy  150  ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                     ||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
```

```
Db    301 FGDVKLLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    319 SPQQLAGKRIGVFSYGSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 SPQQLAGKRIGVFSYGSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
Db    481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 4
US-10-236-417-184
; Sequence 184, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
;   APPLICANT: Agee et al.
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-442C
;   CURRENT APPLICATION NUMBER: US/10/236,417
;   CURRENT FILING DATE: 2003-01-06
;   PRIOR APPLICATION NUMBER: US60/318,120
;   PRIOR FILING DATE: 2001-09-01
;   PRIOR APPLICATION NUMBER: US60/318,430
;   PRIOR FILING DATE: 2001-09-10
;   PRIOR APPLICATION NUMBER: US60/322,781
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/318,184
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,636
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/322,816
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,817
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/361,663
;   PRIOR FILING DATE: 2002-03-05
;   PRIOR APPLICATION NUMBER: US60/396,412
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/323,519
;   PRIOR FILING DATE: 2001-09-19
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 341
;   SOFTWARE: Custom
;   SEQ ID NO 184
;   LENGTH: 520
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-236-417-184

Query Match   98.8%;
Best Local Similarity   91.9%;
  Matches 478; Conservative   0; Mismatches   0; Indels   42; Gaps   1;

                        Score 2480; DB 12; Length 520;
                        Pred. No. 7.3e-238;

```
Qy      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Qy     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Qy    121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------ 149
          |||||||||||||||||||||||||||||
Db    121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAYVATGNARPTGGVGAVALL 180
Qy    150 -------GLRGTMHQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
          |||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSYYCKKI 240
```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKQNSIYSGLEA 258
Db    241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKQNSIYSGLEA 300
Qy    259 FGDVKLLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 FGDVKLLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360
Qy    319 SPQQLAGKRIGVFSYGSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 SPQQLAGKRIGVFSYGSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
Qy    379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWILVRVDEKHRRTYARRPTPNDDTLDEG 480
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
Db    481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 5
US-10-236-417-186
; Sequence 186, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
;   APPLICANT: Agee et al.
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-442C
;   CURRENT APPLICATION NUMBER: US/10/236,417
;   CURRENT FILING DATE: 2003-01-06
;   PRIOR APPLICATION NUMBER: US60/318,120
;   PRIOR FILING DATE: 2001-09-01
;   PRIOR APPLICATION NUMBER: US60/318,430
;   PRIOR FILING DATE: 2001-09-10
;   PRIOR APPLICATION NUMBER: US60/322,781
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/318,184
;   PRIOR FILING DATE: 2001-09-07
;   PRIOR APPLICATION NUMBER: US60/322,636
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/361,663
;   PRIOR FILING DATE: 2002-03-05
;   PRIOR APPLICATION NUMBER: US60/396,412
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/322,816
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,817
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/323,519
;   PRIOR FILING DATE: 2001-09-19
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 341
;   SOFTWARE: Custom
;   SEQ ID NO 186
;   LENGTH: 520
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-236-417-186

Query Match   98.8%;
Best Local Similarity   91.9%;
  Matches 478; Conservative   0; Mismatches   0; Indels   42; Gaps   1;

                        Score 2480; DB 12; Length 520;
                        Pred. No. 7.3e-238;

```
Qy      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Qy     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
```

```
Qy   121  EGIDTTNACYGGTAAVFNAVNWIESSSWD----------------------------  149
          ||||||||||||||||||||||||||||
Db   121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy   150  --------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                        ||||||||||||||||||||||||||||||||||||||||||||
Db   181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy   199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy   259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy   319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLLKSRLDSRTGVAP  378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLLKSRLDSRTGVAP  420

Qy   379  DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy   439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
          |||||||||||||||||||||||||||||||||||||||
Db   481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 6
US-10-236-417-190
Sequence 190, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-442C
CURRENT APPLICATION NUMBER: US/10/236,417
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: US60/318,120
PRIOR FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: US60/318,430
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: US60/322,781
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US60/318,184
PRIOR FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US60/361,663
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US60/396,412
PRIOR FILING DATE: 2002-07-17
PRIOR APPLICATION NUMBER: US60/322,636
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,816
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,817
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/323,519
PRIOR FILING DATE: 2001-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Custom
SEQ ID NO 190
LENGTH: 520
TYPE: PRT
ORGANISM: Homo sapiens
US-10-236-417-190

Query Match   98.8%;
Best Local Similarity   91.9%;
Matches 478;   Conservative 0;

Score 2480;   DB 12;   Length 520;
Pred. No. 7.3e-238;
Mismatches 0;   Indels 42;   Gaps 1;

```
Qy     1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT  60

Qy    61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy   121  EGIDTTNACYGGTAAVFNAVNWIESSSWD----------------------------  149
          ||||||||||||||||||||||||||||
Db   121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy   150  --------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                        ||||||||||||||||||||||||||||||||||||||||||||
Db   181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy   199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy   259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy   319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLLKSRLDSRTGVAP  378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLLKSRLDSRTGVAP  420

Qy   379  DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy   439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
          |||||||||||||||||||||||||||||||||||||||
Db   481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 7
US-10-236-417-194
Sequence 194, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-442C
CURRENT APPLICATION NUMBER: US/10/236,417
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: US60/318,120
PRIOR FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: US60/318,430
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: US60/322,781
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US60/318,184
PRIOR FILING DATE: 2001-09-05
PRIOR APPLICATION NUMBER: US60/361,663
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US60/396,412
PRIOR FILING DATE: 2002-07-17
PRIOR APPLICATION NUMBER: US60/322,636
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,816
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,817
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/323,519
PRIOR FILING DATE: 2001-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Custom
SEQ ID NO 194
LENGTH: 520
TYPE: PRT

```
; ORGANISM: Homo sapiens
US-10-236-417-194

Query Match    98.8%;   Score 2480;   DB 12;   Length 520;
Best Local Similarity  91.9%;   Pred. No. 7.3e-238;
Matches 478;  Conservative  0;  Mismatches  0;  Indels  42;  Gaps  1;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD---------------------           149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy  150  ---------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                  |||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      478
         |||||||||||||||||||||||||||||||||||||||
Db  481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      520
```

```
Query Match    98.8%;   Score 2480;   DB 12;   Length 520;
Best Local Similarity  91.9%;   Pred. No. 7.3e-238;
Matches 478;  Conservative  0;  Mismatches  0;  Indels  42;  Gaps  1;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD---------------------           149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy  150  ---------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                  |||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      478
         |||||||||||||||||||||||||||||||||||||||
Db  481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      520
```

Left column (RESULT 10)

PRIOR APPLICATION NUMBER: US60/322,636
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,817
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,816
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,519
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/323,519
PRIOR FILING DATE: 2001-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Custom
SEQ ID NO 202
TYPE: PRT
LENGTH: 520
ORGANISM: Homo sapiens
US-10-236-417-202

Query Match 98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

```
Qy     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQQAKMGFCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQQAKMGFCT 60

Qy    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNAVWIESSSWD--------------------- 149
         ||||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNAVWIESSSWDGRYALVVAGDIAVVATGNARPTGGVGAVALL 180

Qy   150 ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                   ||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQY 318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQQLAGKRIGVFSYGSGLAAATLYSLKVTQDATPGGSALDKITASLCDLKSRLDSRTGVAP 378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQQLAGKRIGVFSYGSGLAAATLYSLKVTQDATPGGSALDKITASLCDLKSRLDSRTGVAP 420

Qy   379 DVFAENMKLREDTHHLVANYIPQGSISDSLFEGTWYLVRVDEKHRRTYARRPTNDDTLDEG 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLVANYIPQGSISDSLFEGTWYLVRVDEKHRRTYARRPTNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
         ||||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 10
US-10-236-417-204
Sequence 204, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-442C
CURRENT APPLICATION NUMBER: US/10/236,417
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: US60/318,120
PRIOR FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: US60/318,430
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: US60/322,781

Right column (RESULT 11)

PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/318,184
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US60/361,663
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US60/396,412
PRIOR FILING DATE: 2002-07-17
PRIOR APPLICATION NUMBER: US60/322,636
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,817
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,816
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/323,519
PRIOR FILING DATE: 2001-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Custom
SEQ ID NO 204
TYPE: PRT
LENGTH: 520
ORGANISM: Homo sapiens
US-10-236-417-204

Query Match 98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity 91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

```
Qy     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQQAKMGFCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVVDQAELEKYDGVDAGKYTIGLGQQAKMGFCT 60

Qy    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNAVWIESSSWD--------------------- 149
         ||||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNAVWIESSSWDGRYALVVAGDIAVVATGNARPTGGVGAVALL 180

Qy   150 ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                   ||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQY 318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQQLAGKRIGVFSYGSGLAAATLYSLKVTQDATPGGSALDKITASLCDLKSRLDSRTGVAP 378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQQLAGKRIGVFSYGSGLAAATLYSLKVTQDATPGGSALDKITASLCDLKSRLDSRTGVAP 420

Qy   379 DVFAENMKLREDTHHLVANYIPQGSISDSLFEGTWYLVRVDEKHRRTYARRPTNDDTLDEG 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLVANYIPQGSISDSLFEGTWYLVRVDEKHRRTYARRPTNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
         ||||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 11
US-10-236-417-206
Sequence 206, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-442C

RESULT 12

```
;  CURRENT APPLICATION NUMBER: US/10/236,417
;  CURRENT FILING DATE: 2003-01-06
;  PRIOR APPLICATION NUMBER: US60/318,120
;  PRIOR FILING DATE: 2001-09-01
;  PRIOR APPLICATION NUMBER: US60/318,430
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: US60/322,781
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/318,184
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,636
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,817
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US60/361,663
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US60/396,412
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: US60/322,816
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/323,519
;  PRIOR FILING DATE: 2001-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom
;  SEQ ID NO 206
;  LENGTH: 520
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-236-417-206

Query Match             98.8%;  Score 2480;  DB 12;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 7.3e-238;
Matches  478;  Conservative    0;  Mismatches    0;  Indels   42;  Gaps    1;

Qy      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy    121  EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------  149
           ||||||||||||||||||||||||||||||
Db    121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVVAGDIAVYATGNARPTGGVGAVALL  180

Qy    150  ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI   198
                     ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy    199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy    259  FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy    319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy    379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy    439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      478
           |||||||||||||||||||||||||||||||||||||||
Db    481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      520
```

```
US-10-236-417-208
;  Sequence 208, Application US/10236417
;  Publication No. US20040048256A1
;  GENERAL INFORMATION:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-442C
;  CURRENT APPLICATION NUMBER: US/10/236,417
;  CURRENT FILING DATE: 2003-01-06
;  PRIOR APPLICATION NUMBER: US60/318,120
;  PRIOR FILING DATE: 2001-09-01
;  PRIOR APPLICATION NUMBER: US60/318,430
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: US60/322,781
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/318,184
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,636
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,817
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US60/361,663
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US60/396,412
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: US60/322,816
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/323,519
;  PRIOR FILING DATE: 2001-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom
;  SEQ ID NO 208
;  LENGTH: 520
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-236-417-208

Query Match             98.8%;  Score 2480;  DB 12;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 7.3e-238;
Matches  478;  Conservative    0;  Mismatches    0;  Indels   42;  Gaps    1;

Qy      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy    121  EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------  149
           ||||||||||||||||||||||||||||||
Db    121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVVAGDIAVYATGNARPTGGVGAVALL  180

Qy    150  ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI   198
                     ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy    199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy    259  FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy    319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy    379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480
```

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

RESULT 13
US-10-307-817-360
; Sequence 360, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 360
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-817-360

Query Match   98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity   91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative   0; Mismatches   0; Indels 42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------- 149
          ||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYIATGNARPTGGVGAVALL 180

Qy   150 -------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDPGFMIFHSPYCYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDPGFMIFHSPYCYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy   379 DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

RESULT 14
US-10-307-817-370
; Sequence 370, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 370
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-817-370

Query Match   98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity   91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative   0; Mismatches   0; Indels 42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------- 149
          ||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYIATGNARPTGGVGAVALL 180

Qy   150 -------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDPGFMIFHSPYCYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDPGFMIFHSPYCYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy   379 DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLVANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

RESULT 15
US-10-307-817-372
; Sequence 372, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 372
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-307-817-372

Query Match   98.8%; Score 2480; DB 12; Length 520;
Best Local Similarity   91.9%; Pred. No. 7.3e-238;
Matches 478; Conservative   0; Mismatches   0; Indels 42; Gaps   1;

```
Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSWD--------------------------- 149
        |||||||||||||||||||||||||||
Db  121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVVATGNARPTGGVGAVALL 180

Qy  150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                ||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy  259 FGDVKLEDTYFDRDVEKAPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 FGDVKLEDTYFDRDVEKAPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy  379 DVFAENMKLREDTHHLVNYIPQGSSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 DVFAENMKLREDTHHLVNYIPQGSSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
        ||||||||||||||||||||||||||||||||||||||||
Db  481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

Search completed: June 24, 2004, 13:44:07
Job time : 884 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: June 24, 2004, 13:16:22 ; Search time 49 seconds
(without alignments)
503,617 Million cell updates/sec

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_5/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_5/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Searched: 389414 seqs, 51625971 residues

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Sequence: 1 MPGSLPLNAEACWPKDVGIV..............PRLPATAAEPEAAVISNGEH 478

Title: US-10-622-516-2
Perfect score: 2511

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2511	100.0	478	4	US-09-819-993-2	Sequence 2, Appli
2	2511	100.0	478	4	US-10-193-295-2	Sequence 2, Appli
3	2480	98.8	520	4	US-09-819-999-4	Sequence 4, Appli
4	2480	98.8	520	4	US-10-193-295-4	Sequence 4, Appli
5	2439	97.1	518	4	US-09-819-993-5	Sequence 5, Appli
6	2439	97.1	518	4	US-10-193-295-5	Sequence 5, Appli
7	2439	97.1	520	4	US-08-305-505-5	Sequence 5, Appli
8	2362	94.1	518	4	US-08-305-505-6	Sequence 5, Appli
9	2352	93.7	520	1	US-08-305-505-4	Sequence 4, Appli
10	2097	83.5	522	1	US-08-305-505-2	Sequence 3, Appli
11	1476.5	58.8	507	1	US-08-305-505-3	Sequence 4, Appli
12	912.5	36.3	467	3	US-09-306-59C-6	Sequence 6, Appli
13	912.5	36.3	467	3	US-08-925-388-6	Sequence 6, Appli
14	686	27.3	168	4	US-09-401-064-198	Sequence 198, App
15	315	12.5	388	4	US-09-107-532A-5369	Sequence 5369, Ap
16	284.5	11.3	436	4	US-09-134-001C-4295	Sequence 4295, Ap
17	282	11.2	449	4	US-09-134-000C-5444	Sequence 5444, Ap
18	209.5	8.3	316	4	US-09-241-750-2	Sequence 2, Appli
19	123.5	4.9	1495	4	US-09-543-681A-5986	Sequence 5986, Ap
20	123	4.9	358	4	US-09-543-681A-8134	Sequence 8134, Ap
21	105.5	4.2	1143	2	US-09-301-085-108	Sequence 108, App
22	105.5	4.2	1143	4	US-08-310-912A-108	Sequence 108, App
23	105.5	4.2	1143	4	PCT-US95-04589-108	Sequence 108, App
24	105.5	4.2	1144	5	US-08-261-663A-2	Sequence 2, Appli
25	105.5	4.2	1144	1	US-09-357-206A-3	Sequence 3, Appli
26	105.5	4.2	1144	4	US-09-813-742A-3	Sequence 3, Appli
27	105.5	4.2	1144	4	US-09-813-742A-3	
28	105.5	4.2	1144	5	PCT-US95-0775A4-2	Sequence 2, Appli
29	105.5	4.2	1144	5	PCT-US95-0775A4-4	Sequence 4, Appli
30	105	4.2	1494	3	US-08-755-587-186	Sequence 186, App
31	103.5	4.1	3169	4	US-09-453-702B-257	Sequence 257, App
32	103	4.1	1252	4	US-10-012-762-20	Sequence 20, Appl
33	103	4.1	1252	4	US-09-704-036B-20	Sequence 20, Appl
34	102.5	4.1	1144	3	US-08-930-996A-9	Sequence 9, Appl
35	102	4.1	784	3	US-08-846-234-5	Sequence 5, Appli
36	99.5	4.0	652	1	US-08-261-663A-6	Sequence 6, Appli
37	99.5	4.0	652	1	US-09-357-206A-5	Sequence 5, Appli
38	99.5	4.0	652	4	US-09-813-742A-5	Sequence 5, Appli
39	99.5	4.0	876	5	PCT-US95-0775A4-6	Sequence 6, Appli
40	99.5	4.0	876	3	US-08-785-429-2	Sequence 2, Appli
41	99.5	4.0	876	4	US-08-996-621-2	Sequence 2, Appli
42	96.5	3.8	616	4	US-09-236-063-1	Sequence 1, Appl
43	96.5	3.8	640	4	US-09-351-814-13	Sequence 13, Appl
44	96.5	3.8	3248	1	US-08-353-700-1	Sequence 1, Appl
45	96.5	3.8	3248	5	PCT-US95-16216-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-819-993-2
; Sequence 2, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-2

Query Match 100.0%; Score 2511; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy 121 EGIDTTNACYGGTRAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 EGIDTTNACYGGTRAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180

Qy 181 QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND 240
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND 240

Qy 241 FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300

Qy 301 NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKIT 360
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKIT 360

Qy 361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLLVNYIPQGSSIDSLFEGTWYLVRVDEKH 420
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2
US-10-193-295-2
; Sequence 2, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00119SDIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-2

Query Match   100.0%; Score 2511; DB 4; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVYDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVYDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180
Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI 180

Qy 181 QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND 240
Db 181 QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND 240

Qy 241 FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300
Db 241 FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG 300

Qy 301 NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKIT 360
Db 301 NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKIT 360

Qy 361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420
Db 361 ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH 420

Qy 421 RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db 421 RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478

RESULT 3
US-09-819-993-4
; Sequence 4, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC

RESULT 4
US-10-193-295-4
; Sequence 4, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00119SDIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 520
; TYPE: PRT

Query Match   98.8%; Score 2480; DB 4; Length 520;
Best Local Similarity 91.9%; Pred. No. 3.1e-242;
Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVYDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVYDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Db 61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy 121 EGIDTTNACYGGTAAVFNAVNWIESSSWD---------------------------- 149
Db 121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVATGNARAPTGGGVGAVALL 180

Qy 150 ------------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
Db 181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy 199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Db 241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy 259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
Db 301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy 319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Db 361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy 379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
Db 421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy 439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
Db 481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Human
US-09-819-993-4

; ORGANISM: Human
US-10-193-295-4

Query Match   98.8%;   Score 2480; DB 4; Length 520;
Best Local Similarity   91.9%;   Pred. No. 3.1e-242;
Matches 478; Conservative   0; Mismatches   0; Indels   42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------------- 149
          |||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVAYATGNARPTGGVGAVALL 180

Qy   150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                  |||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy   379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKVPRLPATAAEPEAAVISNGEH 478
          |||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKVPRLPATAAEPEAAVISNGEH 520

RESULT 5
US-09-819-993-5
; Sequence 5, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human

US-09-819-993-5

Query Match   97.1%;   Score 2439; DB 4; Length 518;
Best Local Similarity   90.9%;   Pred. No. 4.3e-238;
Matches 471; Conservative   1; Mismatches   4; Indels   42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------------- 149

Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVAYATGNARPTGGVGAVALL 180

Db   150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198

Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Db   199 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLKA 258

Db   241 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Db   319 SPQHLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAQ 420

Db   361 SPQHLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAQ 420

Db   379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Db   439 VGLVHSNIATEHIPSPAKVPRLPATAAEPEAAVISNG 476
          ||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKVPRLPATAAEPEAAVISNG 518

RESULT 6
US-10-193-295-5
; Sequence 5, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195DIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human

US-10-193-295-5

Query Match   97.1%;   Score 2439; DB 4; Length 518;
Best Local Similarity   90.9%;   Pred. No. 4.3e-238;
Matches 471; Conservative   1; Mismatches   4; Indels   42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------------- 149
          |||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVAYATGNARPTGGVGAVALL 180

Qy   150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                  |||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

RESULT 7
US-08-305-505-6
; Sequence 6, Application US/08305505
; Patent No. 5668001
GENERAL INFORMATION:
  APPLICANT: Miziorko, Henry M.
  TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
  TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
  TITLE OF INVENTION: STABILITY
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/305,505
    FILING DATE:
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/08/072,040
    FILING DATE: 02 JUNE 1993
  ATTORNEY/AGENT INFORMATION:
    NAME: Baker, Jean C.
    REGISTRATION NUMBER: 35,433
    REFERENCE/DOCKET NUMBER: 65-053-9083-9
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (414) 277-5591
    TELEFAX: (414) 277-5709
  INFORMATION FOR SEQ ID NO: 6:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 520 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein

US-08-305-505-6

  Query Match   97.1%; Score 2439; DB 1; Length 520;
  Best Local Similarity 90.9%; Pred. No. 4.4e-238;
  Matches 471; Conservative   1; Mismatches   4; Indels   42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
Qy   241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLKA 300

Db   259 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
Qy   301 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Db   319 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
Qy   361 SPQHLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAQ 420

Db   379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
Qy   421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Db   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 476
Qy   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 518

RESULT 8
US-08-305-505-5
; Sequence 5, Application US/08305505
; Patent No. 5668001
GENERAL INFORMATION:
  APPLICANT: Miziorko, Henry M.
  TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
  TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
  TITLE OF INVENTION: STABILITY
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/305,505
    FILING DATE:
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/08/072,040
    FILING DATE: 02 JUNE 1993
  ATTORNEY/AGENT INFORMATION:
    NAME: Baker, Jean C.
    REGISTRATION NUMBER: 35,433
    REFERENCE/DOCKET NUMBER: 65-053-9083-9
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (414) 277-5591
    TELEFAX: (414) 277-5709
  INFORMATION FOR SEQ ID NO: 5:
    SEQUENCE CHARACTERISTICS:

Db   1 MPGSLPLNARACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------- 149
          |||||||||||||||||||||||||||

Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVLL 180

Qy   150 ----------GLRCTHMGHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKI 198
                    |||||||||||||||||||||||||||||||||||||||||||||||

Db   181 IGPNAPLIFERGLRGTHMGHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| :|

Db   241 HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLKA 300

Qy   259 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   301 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   361 SPQHLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAQ 420

Qy   379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 476
          |||||||||||||||||||||||||||||||||||||

Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNG 518

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-305-505-5

Query Match     94.1%;  Score 2362;  DB 1;  Length 520;
Best Local Similarity   87.5%;  Pred. No. 2.8e-230;
Matches  455;  Conservative  11;  Mismatches  12;  Indels  42;  Gaps  1;

Qy      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT 60

Qy     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||||||||||||||||||||||||||||||||:||||||||:||||||||||
Db     61 DREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120

Qy    121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------ 149
          ||||||||||||||||||||||||||||||
Db    121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYASGNARPTGGVGAVALL 180

Qy    150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                  |||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IGPNAPVIFDRGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy    199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300

Qy    259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db    301 FGDVKLEDTYFDRDVEKAFMKASAELFNQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy    319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db    361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP 420

Qy    379 DVFAENMKLREDTHHLVANYIPQSSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||      :
Db    421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHSLDEG 480

Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          |||||||||||||||||||||||||||||: ||||||
Db    481 VGLVHSNTATEHIPSPAKKVPRLPATSGEPESAVISNGEH 520
```

```
RESULT 9
US-08-305-505-4
Sequence 4, Application US/0830505
Patent No. 5668001
GENERAL INFORMATION:
  APPLICANT: Miziorko, Henry M.
  TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
  TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
  TITLE OF INVENTION: STABILITY
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/305,505
    FILING DATE:
;
```

```
Query Match     93.7%;  Score 2352;  DB 1;  Length 520;
Best Local Similarity   87.5%;  Pred. No. 2.8e-229;
Matches  455;  Conservative  9;  Mismatches  14;  Indels  42;  Gaps  1;

Qy      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT 60

Qy     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          ||||||||||:|||||:|||||:|||||||||||||||||||||:||||||||||||||
Db     61 DREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120

Qy    121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------ 149
          ||||||||||||||||||||||||||||||
Db    121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYATGNARPTGGVGAVALL 180

Qy    150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYRKKI 240

Qy    199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 RAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300

Qy    259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db    301 FGDVKLEDTYFDRDVEKAFMKASAELFNQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy    319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          |||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db    361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKVTASLCDLKSRLDSRTCVAP 420

Qy    379 DVFAENMKLREDTHHLVANYIPQSSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||:||||||||||||||||||||||||||||  ||  :
Db    421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHNLGDG 480

Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          |||||||||||||||||||||||||||||: ||||||
Db    481 VGLVHSNTATEHIPSPAKKVPRLPATAASESAVISNGEH 520
```

```
RESULT 10
US-08-305-505-2
Sequence 2, Application US/0830505
Patent No. 5668001
GENERAL INFORMATION:
  APPLICANT: Miziorko, Henry M.
  TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
  TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
  TITLE OF INVENTION: STABILITY
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
;
```

```
;   STREET: 411 East Wisconsin Avenue
;   CITY: Milwaukee
;   STATE: Wisconsin
;   COUNTRY: U.S.A.
;   ZIP: 53202
COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/305,505
;   FILING DATE:
PRIOR APPLICATION DATA:
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: US/08/072,040
;   FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
;   NAME: Baker, Jean C.
;   REGISTRATION NUMBER: 35,433
;   REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (414) 277-5709
;   TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 522 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-305-505-2

Query Match     83.5%;  Score 2097;  DB 1;  Length 522;
Best Local Similarity  76.6%;  Pred. No. 2e-203;
Matches 400;  Conservative 36;  Mismatches 42;  Indels 44;  Gaps 3;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
            |||| |||:|||||||||||||||||||||||| ||||||||||||||||||:|||||||
Db    1  MPGSLPVNTESCWPKDVGIVALEIYFPSQYVDQTELEKYDGVDAGKYTIGLGQSKMGFCS   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
            ||||||||:|||||:||||:||||||||||||||||||||||||||||:|||||||||| :
Db   61  DREDINSLCITVVQKLMERNLSYDCIGRLEVGTETIIDKSKSVKTVLMQLFEESGNTDV   120

Qy  121  EGIDTTNACYGGTAAVFNAVNMIESSSWD-------------------------------  149
            ||||||||||||||| |||:|::|||||||
Db  121  EGIDTTNACYGGTAALFNAINWIESSSWDGRYALVVAGDIAVYATGNARPTGGAGAVAML  180

Qy  150  -----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                       ||||||||||||||||||:|||||:||||||||||||||||||||||:|||
Db  181  VGSNAPLIFERGLRGTHMQHAYDFYKPDMVSEYPVVDGKLSIQCYLSALDRCYSVYRNKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRD-KNSIYGLE   257
            |||||||| : ||||||||||||||||||||||||:|| ||:|||::||   :||:| ||
Db  241  HAQWQKEGTDRGFTLNDFGFMIFHSPYCKLVQKSVARLLINDFLSDQNAETANGVFSGLE  300

Qy  258  AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQ   317
            || |||||| ||||||||||||||||:|||||||||||||||||||||:|||:||:|||||
Db  301  AFRDVKLEDIYFDRDVEKAFMKASAELFNQKTKASLLVSNQGNMYTPSVYGCLASLLAQ   360

Qy  318  YSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSSALDKITASLCDLKSRLDSRTGVA 377
            |||::||| || : ||| |: |||||||:|::|||||||||||||||||||| :||:||||: ||:
Db  361  YSPEHLAGQRISEFSYGGGFAATLYSIRVTQDATPGSSALDKITASLSDLKARLDSRKCIA  420

Qy  378  PDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDE   437
            |||||||||:|::||||||||: ||:::||:|:|||||||||||||||||||||:  |:: |||:
Db  421  PDVFAENMKIRQETHHLANYIPQCSVEDLFEGTWYLVRVDEKHRRTYARRPVWGDGPLEA   480

Qy  438  GVGLVHSNIATEHIPSPAKKVPRLPATA-AEPEAAVISNGEH   478
            || ||| ::|:|||||||||||||:|||  :  |:|:|||::|
Db  481  GVEVVHPGIVHEHIPSPAKKVPRIPATTESEGSVTVAISNGVH  522
```

```
RESULT 11
US-08-305-505-3
;   Sequence 3, Application US/08305505
;   Patent No. 5668001
GENERAL INFORMATION:
;   APPLICANT: Miziorko, Henry M.
;   TITLE OF INVENTION: 3-HYDROXY-3-METHYLGLUTARYL-CoA
;   TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
;   TITLE OF INVENTION: STABILITY
;   NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Quarles & Brady
;   STREET: 411 East Wisconsin Avenue
;   CITY: Milwaukee
;   STATE: Wisconsin
;   COUNTRY: U.S.A.
;   ZIP: 53202
COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/305,505
;   FILING DATE:
PRIOR APPLICATION DATA:
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: US/08/072,040
;   FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
;   NAME: Baker, Jean C.
;   REGISTRATION NUMBER: 35,433
;   REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (414) 277-5709
;   TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 507 amino acids
;   TYPE: amino acids
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-305-505-3

Query Match     58.8%;  Score 1476.5;  DB 1;  Length 507;
Best Local Similarity  58.9%;  Pred. No. 1.4e-140;
Matches 274;  Conservative 76;  Mismatches 72;  Indels 43;  Gaps 2;

Qy    6  PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI   65
            || ::  ||||||:||||:|||:||||| :|||: :|| |||| |||||: |||||:::|||
Db    1  PLAKDTWPKDVGIIALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI   43

Qy   66  NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT  125
            |||| |||::|||| :| : | ||||||||||||||||:|||:||:|| :||||||||||||
Db   43  NSLCITVVQRLMERKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT  102

Qy  126  TNACYGGTAAVFNAVNMIESSSWD------------------------------------  149
            |||||||||: |||| :|:||||||
Db  103  TNACYGGTASLFNAANWMESSSWDGRYALVVCGDIAVYPSGNRPTGGAGAVAMLIGPKA  162

Qy  150  ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ  203
                  |||||||:::||||||: :||||| |||||||||||||| ||||||:: :::|:::|:
Db  163  PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRRKIQNQWK  222

Qy  204  KEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYGLEAFGDVK   263
            :::|::: ||:||::|||||:|:|:|||||||||:|:|||::::  :| ||:||||::::
Db  223  QAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARLMFNDFLSSSSSDKQNNLYKGLEAFKGIK 342

Qy  264  LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQGNMYTSSVYGSLASVLAQYSPQQL  323
            ||: || ||||:| :||| ::|:||||||| :| :|||||||| | ||| |||:|||:
Db  343  LEETYINKDVDKALLKASLDMFNKKTKASLIYLSTNNGNMYTSSL-GCLASLLSHHSAQEL 401
```

```
Qy  324  AGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVSFAE  383
              ||||| ||||||||| :||  | ||||| ||| :  ||| || |||| ::|
Db  402  AGSRIGAFSYGSGLAASFFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRMSPEEFTE   461

Qy  384  NMKLREDTHHLVNYIPQGSIDSLFEGTWLYLVRVDEKHRRTYARRP  428
              :|| ||  |: :| ||| ::|| ||||| |||||| |||:||||
Db  462  IMNQREQYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRKYARRP   506
```

RESULT 12
US-09-306-595C-6
; Sequence 6, Application US/09306595C
; Patent No. 6284506
GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/306,595C
; CURRENT FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 98108210
; PRIOR FILING DATE: 1998-05-06
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 6
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Phaffia rhodozyma

Query Match   36.3%; Score 912.5; DB 3; Length 467;
Best Local Similarity 40.4%; Pred. No. 1.6e-83;
Matches 190; Conservative 81; Mismatches 132; Indels 67; Gaps 6;

```
Qy   14  PKDVGIVALEIYPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV   73
             ||||| |  ||||  :  :| :| ||| ||||||||| :  :|  |||:|||| :||:|
Db   10  PKDVGIILGMEIYPPRRAIAHKDLEAFDGVPSGKYTIGLGNNFMAFTDDTEDINSFALNAV  69

Qy   74  QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNIMQLFEESGNTDIEGIDTTNACYGGT  133
             :| |  :| ||:|||: ||||:|||||||||| ::|||| |||||||||| |||| ||
Db   70  SGLLSKYNVDPKSIGRIDVGTESIIDKSKSVKTVLMDLFESHGNTDIEGIDSKNACYGST  129

Qy  134  AAVFNAVNWIESSSWDG-----------------------------------------LR 152
             ||:||||||||||||||
Db  130  AALFNAVNWIESSSWDGRNAIVFCGDIAIYAEGAARPAGGAGACAILIGPDAPVVFEPVH 189

Qy  153  GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDK---  209
             | : | |:|||| :| |||||||  | :|: |  :|| | |: ||     :| ||:
Db  190  GNFMTNAWDFYKPNLSSEYPIVDGPLSVTSYVNAIDKAVEAYRTKYARPF---GGPKTNG  246

Qy  210  ----------DFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA-F 259
                        :  |: | | |||||| | :| | | :|||| ||   : |:|
Db  247  VTNGHTEVAGVSAASFDYLLFHSPYGKQVVKGHGRLLYNDFRNNPN---DPVFAEVPAEL  303

Qy  260  GDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYS 319
             | :| :   :| :|||  :|:| |  : :| :| :     : | | ||||| |||
Db  304  ATLDMKKSLSDKNVEKSLIAASKSSFNKQVEPGMTTVRQLGNLYTASLFGALASLFSNVP  363

Qy  320  PQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPD  379
             | :|:|||| : :|||| ||: |:||      |:|    :|| |  :||| |:
Db  364  GDELVGKRIALYAYGSGAAASFYALKVKSSTA------FISEKLDLNNRLSNMKIVPCD   416

Qy  380  VFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLYLVRVDEKHRRTYARRPT 429
             :| :  |||: |:|:||  ||||||| | |||| | | |:||::| |||
Db  417  DFVKALKVREETHNAVSYSPIGSLDDLWPGSYYLGEIDSMWRRQYKQVPS  466
```

RESULT 13
US-09-925-388-6
; Sequence 6, Application US/0925388
; Patent No. 6586202
GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/925,388
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US/09/306,595
; PRIOR FILING DATE: 1999-05-06
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 6
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Phaffia rhodozyma

Query Match   36.3%; Score 912.5; DB 4; Length 467;
Best Local Similarity 40.4%; Pred. No. 1.6e-83;
Matches 190; Conservative 81; Mismatches 132; Indels 67; Gaps 6;

```
Qy   14  PKDVGIVALEIYPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV   73
             ||||| |  ||||  :  :| :| ||| ||||||||| :  :|  |||:|||| :||:|
Db   10  PKDVGIILGMEIYPPRRAIAHKDLEAFDGVPSGKYTIGLGNNFMAFTDDTEDINSFALNAV  69

Qy   74  QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNIMQLFEESGNTDIEGIDTTNACYGGT  133
             :| |  :| ||:|||: ||||:|||||||||| ::|||| |||||||||| |||| ||
Db   70  SGLLSKYNVDPKSIGRIDVGTESIIDKSKSVKTVLMDLFESHGNTDIEGIDSKNACYGST  129

Qy  134  AAVFNAVNWIESSSWDG-----------------------------------------LR 152
             ||:||||||||||||||
Db  130  AALFNAVNWIESSSWDGRNAIVFCGDIAIYAEGAARPAGGAGACAILIGPDAPVVFEPVH 189

Qy  153  GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDK---  209
             | : | |:|||| :| |||||||  | :|: |  :|| | |: ||     :| ||:
Db  190  GNFMTNAWDFYKPNLSSEYPIVDGPLSVTSYVNAIDKAVEAYRTKYARPF---GGPKTNG  246

Qy  210  ----------DFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA-F 259
                        :  |: | | |||||| | :| | | :|||| ||   : |:|
Db  247  VTNGHTEVAGVSAASFDYLLFHSPYGKQVVKGHGRLLYNDFRNNPN---DPVFAEVPAEL  303

Qy  260  GDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYS 319
             | :| :   :| :|||  :|:| |  : :| :| :     : | | ||||| |||
Db  304  ATLDMKKSLSDKNVEKSLIAASKSSFNKQVEPGMTTVRQLGNLYTASLFGALASLFSNVP  363

Qy  320  PQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPD  379
             | :|:|||| : :|||| ||: |:||      |:|    :|| |  :||| |:
Db  364  GDELVGKRIALYAYGSGAAASFYALKVKSSTA------FISEKLDLNNRLSNMKIVPCD   416

Qy  380  VFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLYLVRVDEKHRRTYARRPT 429
             :| :  |||: |:|:||  ||||||| | |||| | | |:||::| |||
Db  417  DFVKALKVREETHNAVSYSPIGSLDDLWPGSYYLGEIDSMWRRQYKQVPS  466
```

RESULT 14
US-09-401-064-198
; Sequence 198, Application US/09401064
; Patent No. 6623923
GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Serist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 371

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 198
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-401-064-198

Query Match 100.0%; Score 686; DB 4; Length 168;
Best Local Similarity 27.3%; Pred. No. 2.5e-61;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   39 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 98

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKINLMQLFEESGNTDI 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKINLMQLFEESGNTDI 158

Qy  121 EGIDITTNACY 130
          |||||||||||
Db  159 EGIDITTNACY 168
```

RESULT 15
US-09-107-532A-5369
Sequence 5369, Application US/0910753A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5369:
SEQUENCE CHARACTERISTICS:
LENGTH: 388 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...388
SEQUENCE DESCRIPTION: SEQ ID NO: 5369:

US-09-107-532A-5369

Query Match 12.5%; Score 315; DB 4; Length 388;
Best Local Similarity 24.8%; Pred. No. 4e-23;
Matches 104; Conservative 64; Mismatches 135; Indels 116; Gaps 13;

```
Qy   17 VGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQNL 76
          :   |  | ::|||  | :|||| |||||||||||||||  :|      :  :| :
Db    7 IGIDRLSFPIPNLYLDMTELAESRGDDPAKYHIGIGQDQMAVNRANEDIITLGANAASKI 66

Qy   77 MERNNLSYDCIGRLEVGTETIIDKSKSVKINL------MQLFEESGNTDIEGIDITTNACYG 131
          :       :|||| ||| || |||:  |||       |||||||||||||  |||||
Db   67 VTEKD--RELIDMVIVGTESGIDHSKASAVIIHHLLKIQSFARS------FEVKEACYG 117

Qy  132 GTAAVFNAVNWIE--------- 155
          |||| :|  |:|
Db  118 GTAALHMAKEYVKNHPERKVLIVIASDIARYGLASGGEVTQGVGAVAMMITQNPRILSIED 177

Qy  156 ----MQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKD 210
          :  ||| ||:|| | :|::|::||  :| :|  :|  ||   :::|
Db  178 DSVFLTEDIYDFWRPD-YSEFPVVDGPLSNSTYIESFQKV----------WNRHKELSG 225

Qy  211 FTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFD 270
          : :|  | |:|||| || :  ::| :|  ::     |:  ::|
Db  226 RGLEDYQALIAFHIPYTKMGKKALQSVL------DQTDEEN------------------ 259

Qy  271 RDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLIASVLAQYSPQQLAGKRIGV 330
          :           |||::  :||| ||| |||:||||| ::||:  |||||:|:   ||||
Db  260 ------------QERLMARYESSIRYSRRIGNLYTGSLIYGLTSLL-ENSKSLQPGDRIGL 307

Qy  331 FSYGSSGLAATLYSLKVTQDATPGSALDKITPASLCDLKSRLDSRTGVAPD----VFAENM 385
          |||| |: ||||| |||||: |  ||||:| ||| | ||||||:||||||:    |:|  :
Db  308 FSYGGGAVSEFFTGYLEENYQ------EYLFAQ--SHQEMLDSRTRITVDEYETIFSETL 359
```

Search completed: June 24, 2004, 13:22:33
Job time : 52 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 11:14:21 ; Search time 97 Seconds
                                     (without alignments)
                                     1392.347 Million cell updates/sec

Title: US-10-622-516-2
Perfect Score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV...........PRLPATAAEPEAAVISNGEH 478
Scoring table: BLOSUM62
          Gapop 10.0 , Gapext 0.5
Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
           1: geneseqp1980s:*
           2: geneseqp1990s:*
           3: geneseqp2000e:*
           4: geneseqp2001s:*
           5: geneseqp2002e:*
           6: geneseqp2003ae:*
           7: geneseqp2003bs:*
           8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2511	100.0	478	5	ABG32726	Human hyd
2	2511	100.0	478	6	ABU08379	Human HMG
3	2480	98.8	529	6	ABR54259	Human NOV
4	2480	98.8	520	6	ABR54263	Human NOV
5	2480	98.8	520	6	ABR54261	Human NOV
6	2480	98.8	524	6	ABR54272	Human NOV
7	2480	98.8	520	6	ABR54270	Human NOV
8	2480	98.8	520	6	ABR54268	Human NOV
9	2480	98.8	520	6	ABR54269	Human NOV
10	2480	98.8	520	6	ABR54258	Human NOV
11	2480	98.8	520	6	ABR54267	Human NOV
12	2480	98.8	520	6	ABR54257	Human NOV
13	2480	98.8	520	7	ABR54265	Human NOV
14	2480	98.8	520	7	ADE61135	Human Pro
15	2480	98.8	520	7	ADE61139	Human Pro
16	2480	98.8	529	7	ADE38395	Human pro
17	2475	98.6	520	6	ABR54262	Human NOV
18	2475	98.6	524	6	ABR54264	Human NOV
19	2475	98.6	524	6	ABR54272	Human NOV
20	2475	98.6	525	6	ABR54271	Human NOV
21	2475	98.6	525	6	ABR54266	Human NOV
22	2439	97.1	520	7	ABM04791	Human src
23	2362	94.1	520	7	ADD14146	Human src
24	2362	94.1	520	7	ADE61133	Rat Prote
25	2362	94.1	520	7	ADE61137	Rat Prote
26	2097	83.5	522	2	AAW32222	Avian 3-h
27	1737	69.2	346	4	ABG20556	Novel hum
28	1513	60.3	508	4	AAM78869	Human pro
29	1513	60.3	508	4	ABR47486	Breast ca
30	1513	60.3	508	4	ADB75350	Prostate
31	1513	60.3	508	7	ADE57790	Human Pro
32	1513	60.3	508	7	ADE57794	Human Pro
33	1513	60.3	518	4	AAM79953	Human pro
34	1510	60.1	518	7	ADE09416	Novel pro
35	1494	59.5	508	7	ADE57792	Rat Prote
36	1494	59.5	508	7	ADE57788	Rat Prote
37	1337	53.3	465	4	ABB66034	Drosophil
38	1337	53.2	465	4	ABB60545	Drosophil
39	1336	53.2	583	7	ADE08499	Novel pro
40	1286.5	51.2	453	6	AAE38232	Human enz
41	1016.5	40.5	460	6	ABJ26073	Aspergill
42	1013	40.3	464	6	AAO27022	Acetyl-co
43	982	39.1	461	6	AAO27022	Acetyl-co
44	981.5	39.1	476	6	ABU62384	Rice hydr
45	980	39.0	461	3	AAG43333	Arabidops

ALIGNMENTS

RESULT 1
ABG32726
ID   ABG32726 standard; protein; 478 AA.
XX
AC   ABG32726;
XX
DT   22-NOV-2002 (first entry)
XX
DE   Human hydroxymethylglutaryl-coenzyme A synthase protein.
XX
KW   Human; enzyme; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;
KW   HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;
KW   antibody; synthase; carcinoma.
XX
OS   Homo sapiens.
XX
XX
PN   US6436692-B1.
XX
PD   20-AUG-2002.
XX
PF   29-MAR-2001; 2001US-00819993.
XX
PR   29-MAR-2001; 2001US-00819993.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-689940/74.
DR   N-PSDB; ABS54409.
XX
PT   New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA
PT   synthase, useful as model for the development of human therapeutic
PT   targets and for identifying therapeutic proteins.
XX
PS   Claim 4; Fig 2; 62pp; English.
XX
CC   The invention discloses an isolated nucleic acid molecule encoding a
CC   human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC   important in cholesterologenesis. The polynucleotide and polypeptide are
CC   useful as models for the development of human therapeutic targets, to aid
CC   in the identification of therapeutic proteins and as targets for the
CC   development of human therapeutic agents that modulate the activity of the
CC   polypeptide in cells and tissues. The polynucleotide is useful for
CC   monitoring the effectiveness of modulating compounds on the expression or
CC   activity of the enzyme gene in clinical trials and in treatments; in
CC   diagnostic assays for qualitative changes in expression of enzyme nucleic
CC   acid, to detect mutations in enzyme genes and gene expression products,

```
CC   such a mRNA, for testing an individual for a genotype and as a diagnostic
CC   target that can be used to tailor treatment in an individual. The
CC   polypeptide is useful to raise antibodies, as a reagent in assays
CC   designed to quantitatively determine levels of the protein in biological
CC   fluids, as markers for tissues in which the corresponding protein is
CC   preferentially expressed (e.g. carcinomas), in biological assays related
CC   to members of the synthase subfamily, in drug screening assays, in
CC   competition binding assays, in cell-based or cell-free systems, in
CC   pharmacogenomic analysis and for treating a disorder characterised by an
CC   absence of, inappropriate, or unwanted expression of the polypeptide. The
CC   sequence presented is the human HMG-CoA synthase protein, the gene for
CC   which is located on chromosome 5
XX
SQ   Sequence 478 AA;

Query Match        100.0%;  Score 2511;  DB 5;  Length 478;
Best Local Similarity    100.0%;  Pred. No. 6.4e-237;
Matches  478;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEIIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEIIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180

Qy  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240

Qy  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300

Qy  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKIT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKIT 360

Qy  361  ASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH  420

Qy  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
```

```
RESULT 2
ABU08379
ID   ABU08379 standard; protein; 478 AA.
XX
AC   ABU08379;
XX
DT   05-JUN-2003 (first entry)
XX
DE   Human HMG-CoA synthase-like enzyme.
XX
KW   Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KW   HMG-CoA synthase; immune response; drug screening assay;
KW   pharmacogenomic analysis; enzyme.
XX
OS   Homo sapiens.
XX
PN   US2002173018-A1.
XX
PD   21-NOV-2002.
XX
PF   12-JUL-2002; 2002US-00193295.
```

```
XX
PR   29-MAR-2001; 2001US-00819993.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR   WPI; 2003-352594/33.
DR   N-PSDB; ABX93299, ABX93300.
XX

Claim 1; Fig 2A; 65pp; English.
```

```
New isolated human synthase peptide and gene encoding the enzyme, useful
as models for developing human therapeutic targets, aid in the
identification of therapeutic proteins and as therapeutic targets.
```

```
The present invention relates to the isolation of a novel human enzyme
that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also
discloses polynucleotide sequences encoding the novel enzyme of the
invention. Both the polypeptide and polynucleotide sequences are useful
as models for the development of human therapeutics, for identifying
therapeutic proteins, as targets for development of human therapeutic
agents, and as query sequences to perform a search against sequence data
bases to identify other family members of related sequences. The
polypeptide is useful to raise antibodies or to elicit another immune
response, as a reagent in assays designed to quantitatively determine
levels of the protein in biological fluids, as markers for tissues in
which the corresponding protein is preferentially expressed, in drug
screening assays, in cell-based or cell-free systems, to identify
compounds that modulate synthase activity of the protein in its natural
state, or an altered form that causes the specific disease or pathology
associated with the synthase, to screen a compound for the ability to
stimulate or inhibit interaction with the synthase, to screen a
molecule that normally interacts with the synthase protein, and in
pharmacogenomic analysis. The polynucleotide is useful for monitoring the
effectiveness of modulating compounds on the expression or activity of
the human synthase gene in clinical trials or in a treatment regimen, in
diagnostic assays for qualitative changes in a human synthase nucleic
acid that leads to a pathology, for testing an individual for a genotype
that while not necessarily causing a disease, nevertheless affects the
treatment modality, and as antisense constructs to control human synthase
gene expression in cells, tissues and organisms. The present sequence
represents human HMG-CoA synthase-like enzyme
XX
SQ   Sequence 478 AA;

Query Match        100.0%;  Score 2511;  DB 6;  Length 478;
Best Local Similarity    100.0%;  Pred. No. 6.4e-237;
Matches  478;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEIIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTEIIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSI  180

Qy  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  QCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLND  240

Qy  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  FLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNG  300

Qy  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKIT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKIT 360
```

```
Db  301  NMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKIT  360
Qy  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH   420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKH   420
Qy  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH    478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  RRTYARRPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH    478
```

AC   ABR54259;
XX
ID   ABR54259 standard; protein; 520 AA.

RESULT 3
ABR54259
DT   23-JUN-2003 (first entry)
DE   Human NOV41c protein SEQ ID NO:186.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi C, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, TwomLow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62328.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 1; Page 241; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, anorexia, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;

  Query Match   98.8%; Score 2480; DB 6; Length 520;
  Best Local Similarity   91.9%;   Pred. No. 8.1e-234;
  Matches 478; Conservative 0; Mismatches 0; Indels 42; Gaps 1;

```
Qy    1  MPGSLPLNAEACWPRDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKINLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------  149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYAYGNARPTGGVGAVALL  180

Qy  150  ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI      198
               ||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGPMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEGNDKDFTLNDFGPMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319  SPQQLAGKRIGVFSYGSGSLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWIVLVRVDEKHRRTYARRPTPNDDTLDEG 438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Qy  421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWIVLVRVDEKHRRTYARRPIPNDDTLDEG 480

Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
            ||||||||||||||||||||||||||||||||||||||||
Db  481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 4

ABRS4263

ID   ABRS4263 standard; protein; 520 AA.
XX
AC   ABRS4263;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV41g protein SEQ ID NO.194.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0327811P.
PR   17-SEP-2001; 2001US-0328168P.
PR   17-SEP-2001; 2001US-0328167P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   25-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359999P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX

PI   Agee ML,   Alsobrook JP, Anderson DW, Berghs C,   Boldog FL;
PI   Burgess CE,   Casman SJ,   Catterton E, Chant JS,   Chaudhuri A;
PI   Crabtree J,   Dipippo VA,   Edinger SR,   Eisen AJ,   Ellerman K;
PI   Gangolli EA, Gerlach VL,   Giot L,   Gorman L,   Guo X, Gusev VY, Ji W;
PI   Kekuda R,   Khramtsov NV, Leach MD,   Lepley DM,   Li L,   Liu X;
PI   Malyankar UM, Miller CE,   Ooi CE,   Ott T,   Padigaru M, Patturajan M;
PI   Pena CEA,   Rieger DK,   Rothenberg ME, Shenoy SG,   Shimkets RA;
PI   Spaderna SK, Spytek KA,   Taupier RJ,   Twomlow N,   Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62332.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 1; Page 243; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABRS4167 to ABRS4276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunomodulator, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, immunosuppressive, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABRS4277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;

Query Match   98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity   91.9%;
Matches   478; Conservative   0; Mismatches   0; Indels   42; Gaps   1;
                                    Pred. No. 8.1e-234;

```
Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------ 149
          |||||||||||||||||||||||||||||
Db  121 EGIDTTNACYGGTAAVFNAVNWIBSSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180

Qy  150 ------GLRGTHMQHAYDEYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
              |||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IGPNAPLIFERGLRGTHMQHAYDEYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVVGSLASVLAQY 318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVVGSLASVLAQY 360

Qy  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
```

```
Db    361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLVRVDEKHRRTYARRPTPNDDTLDEG 438

Dd    421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWLVRVDEKHRRTYARRPTPNDDTLDEG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478

Db    481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

```
RESULT 5

ID    ABR54261
AC    ABR54261;
XX
DT    23-JUN-2003  (first entry)
XX
DE    Human NOV41e protein SEQ ID NO:190.
XX
KW    Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW    anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW    haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW    neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW    gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW    congenital heart defect; aortic stenosis; valve disease; transplantation;
KW    tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW    prostate cancer; neoplasm; lymphoma; uterus cancer;
KW    fertility; haemophilia; hypercoagulation; graft versus host disease;
KW    idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW    Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW    cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW    immune disorder; haematopoietic disorder; dyslipidaemia;
KW    metabolic syndrome X.
OS    Homo sapiens.
XX
PN    WO200302301-A2.
XX
PD    20-MAR-2003.
XX
PF    09-SEP-2002; 2002WO-US028538.
XX
PR    07-SEP-2001; 2001US-0318120P.
PR    07-SEP-2001; 2001US-0318184P.
PR    10-SEP-2001; 2001US-0318430P.
PR    17-SEP-2001; 2001US-0322636P.
PR    17-SEP-2001; 2001US-0322781P.
PR    17-SEP-2001; 2001US-0322816P.
PR    17-SEP-2001; 2001US-0322817P.
PR    19-SEP-2001; 2001US-0323519P.
PR    20-SEP-2001; 2001US-0323631P.
PR    20-SEP-2001; 2001US-0323636P.
PR    25-SEP-2001; 2001US-0324969P.
PR    25-SEP-2001; 2001US-0325091P.
PR    26-SEP-2001; 2001US-0324990P.
PR    14-DEC-2001; 2001US-0341144P.
PR    26-FEB-2002; 2002US-0359599P.
PR    05-MAR-2002; 2002US-0361663P.
PR    03-MAY-2002; 2002US-0377909P.
PR    17-MAY-2002; 2002US-0381483P.
PR    29-MAY-2002; 2002US-0383863P.
PR    02-JUL-2002; 2002US-0393333P.
PR    17-JUL-2002; 2002US-0396412P.
PR    13-AUG-2002; 2002US-0403517P.
PR    06-SEP-2002; 2002US-0023641Z.
XX
PA    (CURA-) CURAGEN CORP.
XX
PI    Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI    Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI    Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
```

```
PI    Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI    Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI    Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI    Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI    Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI    Zerhusen BD, Zhong M;
XX
PS    Claim 1; Page 242; 460pp; English.
XX
PT    Novel human proteins and nucleic acid encoding the proteins, useful for
PT    diagnosis, treatment and prevention of disorders involving the human
PT    protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PR    Claim 1; Page 242; 460pp; English.
XX
CC    The present invention describes isolated human NOVX proteins, where X is
CC    1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC    ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC    hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC    antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC    antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC    antiparkinsonian and antilipaemic activities, and can be used in gene
CC    therapy. NOVX proteins are useful for treating or preventing a pathology
CC    associated with a NOVX protein in humans and for treating a syndrome
CC    associated with the human disease. NOVX nucleic acids, proteins and
CC    antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC    atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC    valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC    congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC    disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC    hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC    disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC    infectious disease, anorexia, cancer-associated cachexia, cancer,
CC    Alzheimer's disease, Parkinson's disease, immune disorders,
CC    haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC    ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC    sequences, which are used in examples given in comparison with
CC    ABR54277 represents a human trypsinogen protein given in the present
CC    the human NOV35b protein in the exemplification of the present invention
XX
SQ    Sequence 520 AA;
```

```
Query Match             98.8%;  Score 2480;  DB 6;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 8.1e-234;
Matches  478;  Conservative  0;  Mismatches  0;  Indels  42;  Gaps  1;

Qy      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy    121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------- 149
            ||||||||||||||||||||||||||||||
Db    121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYAYGNARPTGGVGAVALL 180

Qy    150 ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy    199 HAQWQKEBGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 HAQWQKEBGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

Qy    259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy    319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db
Qy   361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db
Qy   379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db
Qy   421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTVARRPTPNDDTLDEG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DR
Qy   439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
DR
Db   481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

```
RESULT 6
AC   ABR54270
ID   ABR54270 standard; protein; 520 AA.
XX
AC   ABR54270;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human NOV41n protein SEQ ID NO:208.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0324990P.
PR   25-SEP-2001; 2001US-0325091P.
PR   14-DEC-2001; 2001US-0341114P.
PR   26-SEP-2001; 2001US-0341149P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377790P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393322P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML,  Alsobrook JP,  Anderson DW,  Berghs C,  Boldog FL;
PI   Burgess CE,  Casman SJ,  Catterton E,  Chant JS,  Chaudhuri A;
PI   Crabtree J,  Dipippo VA,  Edinger SR,  Eisen AJ,  Ellerman K;
```

```
PI   Gangolli EA,  Gerlach VL,  Giot L,  Gorman L,  Guo X,  Gusev VY,  Ji W;
PI   Kekuda R,  Khramtsov NV,  Leach MD,  Lepley DM,  Li L,  Liu X;
PI   Malvankar UM,  Miller CE,  Ooi CE,  Ort T,  Padigaru M,  Patturajan M;
PI   Pena CEA,  Rieger DK,  Rothenberg ME,  Shenoy SG,  Shimkets RA;
PI   Spaderna SK,  Spytek KA,  Taupier RJ,  Twomlow N,  Vernet CAM,  Voss EZ;
PI   Zerhusen BD,  Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ABC62339.
XX
CC   Claim 1; Page 247; 460pp; English.
XX
PS   Claim 1; Page 247; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating and preventing a pathology
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, scleroderma, congenital heart defects,
CC   aortic stenosis, valve disease, tuberous sclerosis, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;
```

```
Query Match             98.8%;  Score 2480;  DB 6;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 8.1e-234;
Matches  478;  Conservative   0;  Mismatches   0;  Indels  42;  Gaps   1;

Qy     1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT 60

Qy    61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121  EGIDTTNACYGGTAAVFNAVNWIESSSWD---------------------------- 149
          |||||||||||||||||||||||||||||
Db   121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180

Qy   150  ---------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                         ||||||||||||||||||||||||||||||||||||||||||||||
Db   181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    361  SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    379  DVPAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  DVPAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480
```

PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spadera SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62337.
XX

XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune and metabolic disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention with
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;

Query Match   98.8%; Score 2480; DB 6; Length 520;
Best Local Similarity   91.9%; Pred. No. 8.1e-234;
Matches 478; Conservative   0; Mismatches   0; Indels 42; Gaps   1;

Qy   1 MPGSLPLNAERACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAERACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKINLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------ 149
              |||||||||||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYAYGNARPTGGVGAVALL 180

Qy   150 ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                        ||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVVGSLASVLAQY 318
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVVGSLASVLAQY 360

Qy   319 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

---

Db    439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
                |||||||||||||||||||||||||||||||||||||||
Db    481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

---

RESULT 7
ABR54268

ID    ABR54268 standard; protein; 520 AA.
XX
AC    ABR54268;
XX
DT    23-JUN-2003  (first entry)
XX
DE    Human NOV41l protein SEQ ID NO:204.
XX

KW    Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW    anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW    haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW    neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW    gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW    congenital heart defect; aortic stenosis; valve disease; transplantation;
KW    tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW    prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW    fertility; haemophilia; hypercoagulation; graft versus host disease;
KW    idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW    Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW    cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW    immune disorder; haematopoietic disorder; dyslipidaemia;
KW    metabolic syndrome X.
XX
OS    Homo sapiens.
XX
PN    WO2003023001-A2.
XX
PD    20-MAR-2003.
XX
PF    09-SEP-2002;  2002WO-US028538.
XX
PR    07-SEP-2001;  2001US-0318120P.
PR    07-SEP-2001;  2001US-0318184P.
PR    10-SEP-2001;  2001US-0318430P.
PR    17-SEP-2001;  2001US-0322636P.
PR    17-SEP-2001;  2001US-0322781P.
PR    17-SEP-2001;  2001US-0322816P.
PR    17-SEP-2001;  2001US-0322817P.
PR    19-SEP-2001;  2001US-0323519P.
PR    20-SEP-2001;  2001US-0323631P.
PR    20-SEP-2001;  2001US-0323636P.
PR    25-SEP-2001;  2001US-0324969P.
PR    25-SEP-2001;  2001US-0325091P.
PR    26-SEP-2001;  2001US-0325490P.
PR    26-SEP-2001;  2001US-0341144P.
PR    26-FEB-2002;  2002US-0359599P.
PR    05-MAR-2002;  2002US-0361663P.
PR    03-MAY-2002;  2002US-0377908P.
PR    17-MAY-2002;  2002US-0381483P.
PR    29-MAY-2002;  2002US-0383863P.
PR    02-JUL-2002;  2002US-0393332P.
PR    17-JUL-2002;  2002US-0396412P.
PR    13-AUG-2002;  2002US-0403517P.
PR    06-SEP-2002;  2002US-00236417.
XX

PA    (CURA-)  CURAGEN CORP.
XX
PI    Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI    Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI    Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;

```
Db    361 SPQGLAGKRIGVFSYGSGGLAATLYSLKVTQDDATPGSALDKITASLCDLKSRLDSRTGVAP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWVLVRVDEKHRRTYARRPTPNDDTLDEG 438

Db    421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWVLVRVDEKHRRTYARRPTPNDDTLDEG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478

Db    481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
              ||||||||||||||||||||||||||||||||||||||||
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
```

```
RESULT 8
ABR54269

ID     ABR54269 standard; protein; 520 AA.
XX
AC     ABR54269;
XX
DT     23-JUN-2003  (first entry)
XX
DE     Human NOV41m protein SEQ ID NO:206.
XX
KW     Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW     anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW     haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW     neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW     gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW     congenital heart defect; aortic stenosis; valve disease; transplantation;
KW     tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW     prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW     fertility; haemophilia; hypercoagulation; graft versus host disease;
KW     idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW     Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW     cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW     immune disorder; haematopoietic disorder; dyslipidaemia;
KW     metabolic syndrome X.
OS     Homo sapiens.
XX
PN     WO200303001-A2.
XX
PD     20-MAR-2003.
XX
PF     09-SEP-2002;  2002WO-US028538.
XX
PR     07-SEP-2001;  2001US-0318120P.
PR     07-SEP-2001;  2001US-0318184P.
PR     10-SEP-2001;  2001US-0318430P.
PR     17-SEP-2001;  2001US-0322636P.
PR     17-SEP-2001;  2001US-0322781P.
PR     17-SEP-2001;  2001US-0322816P.
PR     17-SEP-2001;  2001US-0322817P.
PR     19-SEP-2001;  2001US-0323519P.
PR     20-SEP-2001;  2001US-0323631P.
PR     20-SEP-2001;  2001US-0323636P.
PR     20-SEP-2001;  2001US-0323636P.
PR     25-SEP-2001;  2001US-0324969P.
PR     25-SEP-2001;  2001US-0325091P.
PR     25-SEP-2001;  2001US-0324990P.
PR     14-DEC-2001;  2001US-0341144P.
PR     26-FEB-2002;  2002US-0355599P.
PR     05-MAR-2002;  2002US-0316663P.
PR     03-MAY-2002;  2002US-0377908P.
PR     17-MAY-2002;  2002US-0381483P.
PR     29-MAY-2002;  2002US-0383843P.
PR     02-JUL-2002;  2002US-0393332P.
PR     17-JUL-2002;  2002US-0396412P.
PR     13-AUG-2002;  2002US-0403517P.
PR     06-SEP-2002;  2002US-00256417.
XX
XX
PA     (CURA-)  CURAGEN CORP.
XX
PI     Agee ML,  Alsobrook JP,  Anderson DW,  Berghs C,  Boldog FL;
PI     Burgess CE,  Casman SJ,  Catterton E,  Chant JS,  Chaudhuri A;
PI     Crabtree J,  DiPippo VA,  Edinger SR,  Eisen AJ,  Ellerman K;
```

```
PI     Gangolli EA,  Gerlach VL,  Glot L,  Gorman L,  Guo X,  Gusev VY,  Ji W;
PI     Kekuda R,  Khramtsov NV,  Leach MD,  Lepley DM,  Li L,  Liu X;
PI     Malyankar UM,  Miller CE,  Ooi CE,  Ort T,  Padigaru M,  Patturajan M;
PI     Pena CEA,  Rieger DK,  Rothenberg ME,  Shenoy SG,  Shinkets RA;
PI     Spaderna SK,  Spytek KA,  Taupier RJ,  Twomlow N,  Vernet CAM,  Voss EZ;
PI     Zerhusen BD,  Zhong M;
XX
DR     WPI; 2003-313241/30.
DR     N-PSDB; ACC62338.
XX
PT     Novel human proteins and nucleic acid encoding the proteins, useful for
PT     diagnosis, treatment and prevention of disorders involving the human
PT     protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS     Claim 1; Page 246; 460pp; English.
XX
CC     The present invention describes isolated human NOVX proteins, where X is
CC     1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC     ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC     hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC     antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC     antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC     antiparkinsonian and antilipaemic activities, and can be used in gene
CC     therapy. NOVX proteins are useful for treating or preventing a pathology
CC     associated with a NOVX protein in humans and for treating a syndrome
CC     associated with the human disease. NOVX nucleic acids, proteins and
CC     antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC     atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC     valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC     congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC     disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC     hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC     disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC     infectious disease, anorexia, cancer-associated cachexia, cancer,
CC     Alzheimer's disease, Parkinson's disease, immune disorders,
CC     haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC     ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC     sequences, which are used in examples from the present invention.
CC     ABR54277 represents a human trypsinogen protein given in comparison with
CC     the human NOV35b protein in the exemplification of the present invention
XX
SQ     Sequence 520 AA;

Query Match             98.8%;   Score 2480;  DB 6;  Length 520;
Best Local Similarity   91.9%;   Pred. No. 8.1e-234;
Matches  478;  Conservative  0;  Mismatches  0;  Indels  42;  Gaps  1;
```

```
Qy     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDITTNACYGGTAAVFNAVNWIESSSWD----------------------------- 149
           |||||||||||||||||||||||||||||
Db   121 EGIDITTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVVATGNARPTGGVGAVALL 180

Qy   150 --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                   ||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 FGDVKLEDTYFPDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQGLAGKRIGVFSYGSGGLAATLYSLKVTQDDATPGSALDKITASLCDLKSRLDSRTGVAP 378
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 9

ID   ABR54258 standard; protein; 520 AA.
XX
AC   ABR54258;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human NOV41b protein SEQ ID NO:184.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323531P.
PR   20-SEP-2001; 2001US-0323536P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton B, Chant JS, Chaudhuri A;
PI   Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ott T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62327.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 1; Page 241; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, anorexia, cancer-associated cachexia,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;

Query Match            98.8%;  Score 2480;  DB 6;  Length 520;
Best Local Similarity  91.9%;             Pred. No. 8.1e-234;
Matches  478;  Conservative   0;  Mismatches   0;  Indels  42;  Gaps   1;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGRCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKINLMQLFEESGNTDI 120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------- 149
         ||||||||||||||||||||||||||||                            
Db  121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYAYGNARPTGGVGAVALL 180

Qy  150 -----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                    ||||||||||||||||||||||||||||||||||||||||||||||| 
Db  181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy  319 SPQOLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  361 SPQOLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy  379 DVFAENMKLREDTHHLVNYIPQSSIDSLFEGTWYLVRVDEKHRRTYAARPTPNDDTLDEG 438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db  421 DVFAENMKLREDTHHLVNYIPQSSIDSLFEGTWYLVRVDEKHRRTYAARPTPNDDTLDEG 480

Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
         ||||||||||||||||||||||||||||||||||||||| 
Db  481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

```
Db  361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480
Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
        ||||||||||||||||||||||||||||||||||||||||
Db  481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

RESULT 10
ABR54267
```
ID   ABR54267 standard; protein; 520 AA.
XX
AC   ABR54267;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV41k protein SEQ ID NO:202.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiasthmatic; antiinflammatory; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
OS   Homo sapiens.
XX
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0326236P.
PR   17-SEP-2001; 2001US-0327081P.
PR   17-SEP-2001; 2001US-0328162P.
PR   17-SEP-2001; 2001US-0328817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0326311P.
PR   20-SEP-2001; 2001US-0326331P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0249669P.
PR   25-SEP-2001; 2001US-0325091P.
PR   25-SEP-2001; 2001US-0325092P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0059599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   17-MAY-2002; 2002US-0377908P.
PR   03-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   17-MAY-2002; 2002US-0933332P.
PR   02-JUL-2002; 2002US-0933322P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0405172P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VV, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62336.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 1; Page 245; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62336 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiasthmatic, antiinflammatory,
CC   antiasthmatic, metabolic, antiinfertility, immunomodulator, anti-HIV,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62336 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;
```

```
Query Match   98.8%;  Score 2480;  DB 6;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 8.1e-234;
Matches  478;  Conservative   0;  Mismatches   0;  Indels  42;  Gaps   1;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
Qy  121 EGIDTTNACYGGTAAVFNAVVNWIESSSWD------------------------------ 149
        |||||||||||||||||||||||||||||||
Db  121 EGIDTTNACYGGTAAVFNAVVNWIESSSWDGRYALVVAGDIAVATGNARPTGGVGAVALL 180
Qy  150 ----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                  ||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240
Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300
Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360
Qy  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   361  SPQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARPTPNDDTLDEG 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARPTPNDDTLDEG 480
```

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
          ||||||||||||||||||||||||||||||||||||||||
Db   481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520

RESULT 11
ID   ABR54257 standard; protein; 520 AA.
XX
AC   ABR54257;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;

PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Lii L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spadena SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;

Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
Crabtree J, DiPippo VA, Edinger SR, Eisen AJ, Ellerman K;

(CURA-) CURAGEN CORP.

```
Db    361 SPQQLAGKRIGVFSYGSGGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    379 DVFAENMKLREDTHLHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 DVFAENMKLREDTHLHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    439 VGLVHSNIATEHIPSPAKKVPRLPATAAPEAAVISNGEH 478
               |||||||||||||||||||||||||||||||||||||||
Db    481 VGLVHSNIATEHIPSPAKKVPRLPATAAPEAAVISNGEH 520
```

RESULT 12

ABR54265

ID   ABR54265 standard; protein; 520 AA.
XX
AC   ABR54265;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV4ll protein SEQ ID NO:198.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X.
XX
OS   Homo sapiens.
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
```
PR     07-SEP-2001; 2001US-0318120P.
PR     07-SEP-2001; 2001US-0318184P.
PR     10-SEP-2001; 2001US-0318430P.
PR     17-SEP-2001; 2001US-0322636P.
PR     17-SEP-2001; 2001US-0327781P.
PR     17-SEP-2001; 2001US-0322816P.
PR     17-SEP-2001; 2001US-0328817P.
PR     19-SEP-2001; 2001US-0323519P.
PR     20-SEP-2001; 2001US-0323631P.
PR     20-SEP-2001; 2001US-0323636P.
PR     25-SEP-2001; 2001US-0324969P.
PR     25-SEP-2001; 2001US-0325091P.
PR     26-SEP-2001; 2001US-0324990P.
PR     14-DEC-2001; 2001US-0341144P.
PR     26-FEB-2002; 2002US-0359599P.
PR     05-MAR-2002; 2002US-0361663P.
PR     03-MAY-2002; 2002US-0377908P.
PR     17-MAY-2002; 2002US-0381483P.
PR     29-MAY-2002; 2002US-0383863P.
PR     02-JUL-2002; 2002US-0393332P.
PR     17-JUL-2002; 2002US-0396412P.
PR     13-AUG-2002; 2002US-0403512P.
PR     06-SEP-2002; 2002US-0236417.
```
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL,
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
```

PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   N-PSDB; ACC62334.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 1; Page 244; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 520 AA;
```
Query Match            98.8%;  Score 2480;  DB 6;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 8.1e-234;
Matches   478; Conservative  0; Mismatches  0; Indels  42; Gaps   1;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD----------------------------- 149
              |||||||||||||||||||||||||||||
Db  121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180

Qy  150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                         |||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IGFNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTRASLLVSNQNGNMYTSSYVGSLASVLAQY 318
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTRASLLVSNQNGNMYTSSYVGSLASVLAQY 360

Qy  319 SPQQLAGKRIGVFSYGSGGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   361  SPQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYAARPTPNDDTLDEG  438
          |||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db   421  DVFAENMKLREDTHHLVNYIPQSSIDSLFEGTWYLVRVDEKHRRTYARPTPNDDTLDEG

Qy   439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
          ||||||||||||||||||||||||||||||||||||||||
Db   481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  520
```

RESULT 13

```
ID   ADE61135
     ADE61135 standard; protein; 520 AA.
XX
AC   ADE61135;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein Q01581, SEQ ID NO 7053.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C, D'urso D, Befort K, Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; Q01581.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
     preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
```

CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polypeptides or their antibodies. The polynucleotide or the one or more
CC   polypeptides or their antibodies, a method for producing a
CC   pharmaceutical composition, a method for identifying a compound, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The method is useful for treating
CC   pain and spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI) in an animal (e.g. Gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of

```
Qy   Sequence 520 AA;

Qy   Best Local Similarity   98.8%;
Qy   Query Match             91.9%;
     Matches  478;  Conservative   0;  Mismatches   0;  Indels  42;  Gaps   1;
     Score 2480;  DB 7;  Pred. No. 8.1e-234;  Length 520;

Qy     1  MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy    61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
          |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    61  DREDINSLCMTVVQNLMERNNLSYPCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy   121  EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------  149
          ||||||||||||||||||||||||||||
Db   121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVLVVAGDIAVYAYGNARPTGGVGAVALL  180

Qy   150  ------------GLRGTHMQHAYDFVKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKK  198
                      |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  IGPNAPLIFERGLRGTHMQHAYDFVKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKK  240

Qy   199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy   259  FGDVKLEDTYFPDRDVEKAPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  FGDVKLEDTYFPDRDVEKAPMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy   319  SPQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  SPQQLAGKRIGVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy   379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYAARPTPNDDTLDEG  438
          |||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db   421  DVFAENMKLREDTHHLVNYIPQSSIDSLFEGTWYLVRVDEKHRRTYARPTPNDDTLDEG  480

Qy   439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
          ||||||||||||||||||||||||||||||||||||||||
Db   481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  520
```

RESULT 14

```
ID   ADE61139
     ADE61139 standard; protein; 520 AA.
XX
AC   ADE61139;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein Q01581, SEQ ID NO 7057.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
```

CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.

```
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (BARB ) BAYER AG.
XX
PI   Woolf C,  D'urso D,  Befort K,  Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; Q01581.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which did not form part of the printed. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 520 AA;

Query Match             98.8%;  Score 2480;  DB 7;  Length 520;
Best Local Similarity   91.9%;  Pred. No. 8.1e-234;
Matches  478;  Conservative   0;  Mismatches   0;  Indels  42;  Gaps    1;

Qy         1 MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDAGKYYIGLGQARKMGFCT 60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDAGKYYIGLGQARKMGFCT 60

Qy        61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy       121 EGIDTTNACYGGTAAVFNAVNAVNWIESSSWD------------------------- 149
             ||||||||||||||||||||||||||||||
Db       121 EGIDTTNACYGGTAAVFNAVNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGVGAVALL 180

Qy       150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 240

Qy       199 HAQWOKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 HAQWOKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 300

Qy       259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy       319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
```

```
Db       361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 420

Qy       379 DVFAENMKLREDTHHLVNYIPGGSIDSLFEGTWLVRVDEKHRRTYAARPTPNDDTLDEG 438
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 DVFAENMKLREDTHHLVNYIPGGSIDSLFEGTWLVRVDEKHRRTYAARPTPNDDTLDEG 480

Qy       439 VGLVHSNIATEHIPSPAKKVPRLPATAAAEPEAAVISNGEH 478
             |||||||||||||||||||||||||||||||||||||||||
Db       481 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 520
```

```
RESULT 15
ID   ADE38395 standard; protein; 520 AA.
AC   ADE38395;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human protein 9389 amino acid sequence.
XX
KW   tumourigenic disorder; angiogenic disorder; aberrant gene expression;
KW   aberrant protein activity; cytostatic; antithyroid; antidiabetic;
KW   ophthalmological; cancer; breast cancer; colon cancer; lung cancer;
KW   prostatic cancer; Grave's disease; diabetic retinopathy; protein 9389.
OS   Homo sapiens.
XX
PN   WO2003065006-A2.
XX
PD   07-AUG-2003.
XX
PF   30-JAN-2003; 2003WO-US002588.
XX
PR   31-JAN-2002; 2002US-0353600P.
PR   15-MAR-2002; 2002US-0364517P.
PR   09-APR-2002; 2002US-0371075P.
PR   10-APR-2002; 2002US-0371507P.
PR   16-APR-2002; 2002US-0372984P.
PR   19-APR-2002; 2002US-0374194P.
PR   24-MAY-2002; 2002US-0382995P.
PR   31-MAY-2002; 2002US-0385023P.
PR   14-JUN-2002; 2002US-0388853P.
PR   17-JUN-2002; 2002US-0389395P.
PR   25-JUN-2002; 2002US-0391324P.
PR   15-JUL-2002; 2002US-0395944P.
PR   22-JUL-2002; 2002US-0403046P.
PR   13-AUG-2002; 2002US-0403046P.
PR   15-AUG-2002; 2002US-0405155P.
PR   27-AUG-2002; 2002US-0406316P.
PR   09-APR-2002; 2002US-0371075P.
PR   25-OCT-2002; 2002US-0421195P.
PR   12-NOV-2002; 2002US-0425456P.
PR   19-NOV-2002; 2002US-0427626P.
PR   10-DEC-2002; 2002US-0431222P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Hunter JJ, Macbeth KJ, Tsai F,  Lesoon A,  Lightcap ES;
PI   Williamson MW,  Rudolph-Owen LA;
XX
DR   WPI; 2003-646176/61.
DR   N-PSDB; ADE38394.
XX
PS   Disclosure; SEQ ID NO 56; 454pp; English.
XX
PT   Treating subject having tumourigenic disorder or angiogenic disorder
PT   caused by aberrant polypeptide e.g., N-formylpeptide receptor or nucleic
PT   acid, by administering a modulator.
XX
CC   This invention relates to a novel method of treating a human subject
CC   having a tumourigenic disorder or angiogenic disorder, caused by aberrant
CC   gene expression or activity of an isolated protein, by administering a
```

```
CC    modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC    or ophthalmological activity. The method is useful for treating a subject
CC    having a tumourigenic or angiogenic disorder, in particular for treating
CC    cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC    cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC    present sequence is the amino acid sequence of the novel isolated human
CC    protein 9389 of the invention.
XX
SQ    Sequence 520 AA;

Query Match          98.8%;  Score 2480;  DB 7;  Length 520;
Best Local Similarity  91.9%;  Pred. No. 8.1e-234;
Matches 478;  Conservative  0;  Mismatches  0;  Indels  42;  Gaps  1;

Qy      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60

Qy     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy    121  EGIDTTTNACYGGTAAVFNAVNWIESSSWD---------------------------    149
           |||||||||||||||||||||||||||||
Db    121  EGIDTTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVVATGNARPTGSVGAVALL  180

Qy    150  ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                       ||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy    199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  300

Qy    259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy    319  SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy    379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy    439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      478
           |||||||||||||||||||||||||||||||||||||||
Db    481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                      520
```

Search completed: June 24, 2004, 13:17:51
Job time : 101 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 06:24:23 ; Search time 4681 Seconds
(without alignments)
12771.648 Million cell updates/sec

Title: US-10-622-516-1
Perfect score: 2002
Sequence: 1 cgcctcccagcgactctcgg..........aaaaaaaaaaaaaaaaaaa 2002
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 27512289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Maximum DB seq length: 2000000000
Minimum DB seq length: 0

Database : EST:*

```
  1: em_estba:*
  2: em_esthum:*
  3: em_estin:*
  4: em_estmu:*
  5: em_estov:*
  6: em_estpl:*
  7: em_estro:*
  8: em_htc:*
  9: gb_est1:*
 10: gb_est2:*
 11: gb_htc:*
 12: gb_est3:*
 13: gb_est4:*
 14: gb_est5:*
 15: em_estfun:*
 16: em_estom:*
 17: em_gss_hum:*
 18: em_gss_inv:*
 19: em_gss_pln:*
 20: em_gss_vrt:*
 21: em_gss_fun:*
 22: em_gss_mam:*
 23: em_gss_mus:*
 24: em_gss_pro:*
 25: em_gss_rod:*
 26: em_gss_phg:*
 27: em_gss_vrl:*
 28: gb_gss1:*
 29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1191.4	59.5	3222	11	AK045094 AK045094 Mus muscu
2	1189.6	59.4	3286	11	AK044835 AK044835 Mus muscu
3	1189.6	59.4	3466	11	AK031297 AK031297 Mus muscu
4	829.4	41.4	960	12	BQ049460 BQ049460 AGENCOURT

No.	Score		Length	DB	ID	
5	800.0	40.0	881	12	BG697489	BG697489 602660682
6	788.6	39.4	1053	13	BU162936	BU162936 AGENCOURT
7	765.8	38.3	817	12	BG740730	BG740730 602631591
8	737	36.8	738	12	BG742704	BG742704 602632919
9	706.2	35.3	894	12	BG683150	BG683150 602631636
10	666.6	33.3	703	9	AU127534	AU127534
11	664.2	33.2	1083	13	BX441322	BX441322
12	657.6	32.8	795	12	BG683791	BG683791 602651636
13	644.2	32.2	733	12	BG698557	BG698557 602658594
14	643.6	32.1	789	14	CA512392	CA512392 UI-R-FJ0-
15	626	31.3	1201	13	BX419944	BX419944
16	620	31.0	620	12	BM790716	BM790716 K-EST0070
17	618.2	30.9	779	14	CD354310	CD354310 UI-M-GM0-
18	611.8	30.6	780	14	CA316122	CA316122 UI-M-FW0-
19	608.4	30.4	998	12	BM474567	BM474567 AGENCOURT
20	601	30.0	914	10	AW249691	AW249691 2819708.3
21	596.4	29.8	910	12	BU941319	BU941319 AGENCOURT
22	596	29.8	950	14	CF412994	CF412994 CH#083_D
23	595.8	29.8	937	13	BX441044	BX441044
24	592	29.6	592	9	AL138300	AL138300 DKFZp547K
25	580.2	29.0	682	14	CB288167	CB288167 CMD75_F05
26	576	28.8	810	14	CD354238	CD354238 UI-M-GM0-
27	573	28.6	725	14	CD354751	CD354751 UI-M-GM0-
28	572.6	28.6	905	13	BU508388	BU508388 AGENCOURT
29	572.2	28.6	681	14	CF169002	CF169002 B0807G11-
30	568.4	28.4	674	14	CF170711	CF170711 B083ZB04-
31	568.4	28.4	877	9	AU132233	AU132233
32	568.2	28.4	814	9	AU130614	AU130614
33	568	28.4	718	12	BM963690	BM963690 UI-M-BO0-
34	567.4	28.3	760	12	BG771780	BG771780 602720453
35	565.2	28.2	674	14	CA873594	CA873594 K092GC09-
36	561.8	28.1	669	14	CA874751	CA874751 K093AFJ0-
37	560.6	28.0	912	13	BQ720370	BQ720370 603073318
38	558.6	27.9	1059	12	BI463653	BI463653 603207318
39	552	27.6	657	14	CA874680	CA874680 K0934B05-
40	552	27.6	657	14	CA884619	CA884619 K0112ED9-
41	552	27.6	1201	13	BX445244	BX445244 BX445244
42	551.4	27.5	555	10	AW951421	AW951421 EST363491
43	550.4	27.5	656	14	CA889355	CA889355 B0152D07-
44	550.4	27.5	657	14	CA892777	CA892777 B0174F04-
45	550	27.5	654	14	CF900429	CF900429 A0315D11-

ALIGNMENTS

```
RESULT 1
AK045094
LOCUS       AK045094                3222 bp    mRNA    linear   HTC 20-SEP-2003
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130032C06 product:pre B-cell leukemia
            transcription factor 1, full insert sequence.
ACCESSION   AK045094
VERSION     AK045094.1  GI:26337068
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
```

```
REFERENCE   3
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,K., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861
REFERENCE   4  (bases 1 to 3222)
AUTHORS     The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3222)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,J., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saito,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="PANTOM_DB:B130032C06"
                     /db_xref="MGI:2410453"
                     /db_xref="taxon:10090"
                     /clone="B130032C06"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     CDS             62..1624
                     /note="unnamed protein product; pre B-cell leukemia
                     transcription factor 1 [MGD:MGI:97495, GB:NM_008783,
                     evidence: BLASTN, 99%, match=124]
```

putative"
/codon_start=1
/protein_id="BAC32218.1"
/db_xref=GI:26337069"
/translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQQAELEKYDGVDA
GKYILGLGQARMGFCTDREDINSLCLTVVQKLMERHISLSYDICIGRLEVGTETIIDKSK
SVKSNMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
AIYAIGNARPTQGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
KLSIQCYLSALDRCYSVYRKIRAQWQKEGKDKOFTLNDFGFMIFHSPYCKLVQKSLA
RMFLNDFLNDQNRDKNSIYSGLEARGDVKLEDTYFDRDVEKAPMKASSELFNQKTRAS
TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
LVSNQNGNMYISSVYGSSLAVLAQYSPQQLAGHRVGVSYSYSGSGLAATLYSLKVTODA
GTWLVRVDEKHRRTYARRPFTNDHSLDEGMGIJVHSNTATEHIIPSPAKKVPRLPATSA
ESESAVISNGEH"
/note="putative"

polyA_signal   3208..3213
               /note="putative"
polyA_site   3222
               /note="putative"

```
ORIGIN

Query Match                  Best Local Similarity   59.5%;  Score 1191.4;  DB 11;  Length 3222;
Best Local Similarity        79.0%;  Pred. No. 9.7e-175;
Matches 1570; Conservative     0; Mismatches 271; Indels 146; Gaps     7;

Qy        89 TTTCGTGGGTCACTCCCTTTCCTTCTGCTGCCGCTCGGTCACGCTTGCTCTCTTTCACCATGC  148
             ||||  ||||||||||||  |||| || |||| ||||||| ||||||| | || | |||||||
Db         6 TCTCCTTGCTTTGCTCCGTTCTTCTTCCAGGTCGTTCCAGGCCTCGTTCCCGATCCCCTT   65

Qy       149 CTGGATCACTTCCTTTGAATGCAGAAAGCTTGCTGGCCAAAAGAATGTGGGAATTGTTGCCC  208
             |||| |||||||||||||||||| ||| ||||| |||||||||||||| ||||||||| |||
Db        66 CTGGGTCACTTCCTTTGAATGCAGAGGCTTGCTGGCCAAAAGATGTGGGAATCGTTGCCC   125

Qy       209 TTGAGATCTATTTTTCCTTCTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGATGGTG  268
             |||||||||||||||| |||||||||||||||||||||||||||| |||||||| |||||||
Db       126 TTGAAATCTACTTTTCCTTCTCAATATGTCGATCAAGCTGAGTTGGGAAAAATACGATGGTG   185

Qy       269 TAGAGTGCTCGGAAAGTATACAATTGGCTTGGGCAGGCCAAGATAGGGCCGCTTCTGCCACAGATA  328
             |||||||  |||||| ||||||||||||||||||||||||||||||| | ||||||||||||||
Db       186 TAGAGTGCTCGGAAAGTATACCATTCGGCCTGGGCGCCAGGGCCTTCTGCACGGGATC   245

Qy       329 GAGAGAGATATTAAACTTCTTTGCATGACTGTGGTTCAGAAATCTTATGGAGAAAAATAACC  388
             ||||||||||||  |||||||||| |||||||||| |||||||||||||||| |||||||
Db       246 GTGAGAGACATCAACTCTCTTTGCCCTGACTGTGGTTCAGAAAACTGATCGAGAGACATAGCC   305

Qy       389 TTTCCTATGATTGTTGGGCGGCTGGGAGTTGGAACAGAGCAACAATCAATGACAATCAA  448
             |||| |||||||| |||||||||||||||||||||||| ||| ||||||||||||||
Db       306 TTTCCTATGATTGCATTGGGCGGCTAGAAGTTGGACAACAGACAATCAAATCGACAAATCGA   365

Qy       449 AGTCTGTGAAGACTAATTGAATGCAGCGCGTTGAAGAGTTGGAACAGAGATAATAGAAG  508
             ||||||||||||||||||| |||||||||||||||||||||||||||||| |||||||
Db       366 AATCAGTGAAGTCTAATTTGAAGCAGCTGTTTGAGGAGTCGGAATTACAGATAATAGAAG   425

Qy       509 GAATCGACACAACTAATGCATGCATGGGAGGACACAGCTGCTGTTCTTCTCAAATGCTGTTAACT  568
             ||||||| || || ||  |||| |||||||||||| |||||| ||| || | |||| | ||||
Db       426 GAATAGACTACAACCAAATGCATGCTTATGGGGGCACAGCTGCAGTCTTCAAATGCCCTGGAACT   485

Qy       569 GGATTGAGTCCAGCTTCTTGGGAT-----------------------  591
             |||||||| |||| |||||||||||
Db       486 GGGTCGAATCCAGCTTCTTGGGGATGGACCGGATGGATATATGCTCTGGGTAGTTGCAGGAGACATTGCTA   545

Qy       592 -----------------------------------GGGCTTCGTGGGACACATATGCAACATGCCT  622
                                                ||||| || ||||||||||| ||||||||
Db       546 TATATGCCACAGGAAATGCCAGACCTACAGGTGGAGTTGGAGCTGTGGCCCTGGTAATTG   605

Qy       623 ATGATTTTTTACAAGCCGATATTTCTGAATATCTCTATAGTAGTTAGATGGAAACTCTCCA  682
             ||||||  |  ||||||| |  ||||| | |||||| |||||| ||| ||| ||||||||||
Db       606 GGCCAAACGCTCCCCTAATTTTGACCGACGGGCTCCGTGGGGACACACATGCAGCATGCCT   665

Qy       666 ATGACTTTTACAAGCCTGACATGCTCTCCGAGTACCCTGGTGGTTCGACCGGAAAGCTCCCA  725
             ||||||||||  ||| | |||||||| | | || | ||||||| || | ||| ||||||||
Db       666 GAATAGACTACAACCAAATGCATGCTTATGGGGGCACAGCTGCAGTCTTCAAAATCCATG   725

Qy       683 TACAGTGCTACCTCAGTGCATTAGACCGGCTGCTACTCGTTCTGTCTGCAAAAAAGATCCATG  742
```

```
Db  726  TACAGTGCTACCTTCAGCGCCCTGGACCTGCTATTCTTCGTTCTACCGGAAAAGATCCGTG
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Qy  743  CCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTGATTTCATGA  785
Db  786  CCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTGATTTCATGA
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Qy  803  TCTTTCACTCACCATATTGTGTAAACTGGTTCAGAAATCTCTAGCTCCGGATGTTGCTGAATG  802
Db       ...                                                             845
Qy  803  TCTTTCACTCACCATATTGTGTAAACTGGTTCAGAAATCTCTAGCTCCGGATGTTGCTGAATG  862
Db  846  TCTTTCACTCACCATATTGTTGCAGAAATCTCTAGCTCCGGATGTTCCTGAATG          905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  863  ACTTCCTTAATGACCAGAATAGAGAATATAAAAATAGTATGTATCTATAGTGGCCCTTTG   922
Db  906  ACTTTCTTAATGATCAAAACAGAGAATAAAAAACAGTAATTACAGTGGGACTGTTAAGG     965
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  923  GGGATGTTAAAATTAGAGACACCTACTTTGATGAGATGTGGAAGGCATTTATGAAGG      982
Db  966  GGGACGTTAAATTAGAAGTAGTACTTTACTTTGACAGAGATGTAGAAAAGCATTTATGAAGG  1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  983  CTAGCTCTGAACTCTTCAGTCAGAGAAAAACAAAAGCCATCTTTACTTGTATCAAATCAAAATG  1042
Db  1026 CTAGTTCTGAGCTATTCAAACCAGAAAACACAAAGGCGTCTTTCGTCTTTGCTTGTGTCTTAATCAGAATG  1085
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Qy  1043 GAAATATGTACACATCTTCAGTATATAGGTTCCCTTGCATCTGTTCTAGCACAGTACTCAC  1102
Db  1086 GAAATATGTACACATCCTCTTCTGTCTTATGGTTCCTGCTTCTGTTCCTGGCACAGTACTCAC  1145
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1103 CTCAGCAATTACCAGGGAAGAGAAATTGGAGTGGTTGTTTTCTTATGTGTTCTTGGCTGCCA  1162
Db  1146 CTCAGCACTTGGCGAAGAGGGTTGGAGTGGTTCTCTTTACGGTTCTTGGCTGCCA         1205
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1163 CTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCCGGGGTTCTGCTTCTTCTGGGCCA  1222
Db  1206 CGGCTGTACTCCCTTTAAAGTCACACAAGGTACTGCCCTTGATTAACTATATATTCCC     1265
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1223 CAGCAAGTTTATGTGATCTTAAATCAAGGCCTTGATTCAAGAACTGGTGTGTCCAGATG   1282
Db  1266 CAGCAAGTTTATGTGACCTTAAATCAAGGCTTGAACTCGAGAACTTGTGTCCGGATG       1325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1283 TCTTTCGGTCTGAAAAACATGAAGCTCAGAGGAGGACACCCAACATCATTTGGTCAAACTATATATTTCCCC  1342
Db  1326 TCTTTTGCTGAAAAACATGAAGCTCAGAGGAGGCACACAATCACTACTTAGCCAACTATATTCCCC  1385
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1343 AGGGTTCAAATAGATTCACTCCTTTGACTGGTACTACTTAGTTAGGGTGGTGGATGAAAAGC  1402
Db  1386 AGTGTTCAAATAGACTCACTCTTTGAAGGAACGTGGTATCTGGTCAGAGTGGGATGAAAAAAC   1445
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1403 ACAGAAGAACTTAGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGAGTAG     1462
              |||||||||||||||||||||||  |||||||||||||||||||||||
Db  1446 ACAGAAGGACTTACGCCCGCGCCCTTCACAAAATGACCACAGTTTGGATGAAGGAATGG  1505
Db  1506 GGCTCGTCGTCCATAGTAACACAGCAACAGCACATAATTCCAAGCCTGAAGAAGTGCCAA     1565
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1523 GACTCCCCTGCCACAGCAGCAGAACCTGAAGCACAGCACTTAGTTAGTAATGGGAACATTTAAG  1582
Db  1566 GACTCCCCTGCACACCTGCGGAATCTCAGCTGTCGTCATCAGTATGAATAACGTGGGAGCACTGAG  1625
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1583 ATACTCTGTGAGGGTGCAAGACTTCAGGGTGTGGGGTGGGCATGGGGGTGGGGGAACA   1642
              ||| |||||||||||||||||||||||||||||||||||||||||
Db  1626 AGTCTGTTGGCCTTTACA-----------GAGGCTCGGGGCCCGGGGGGGATGGGGTATGGGAAACG  1674
Qy  1643 GTTGGAGGAAATGGGGGATTACTGGGGATTAAATTAAAAGGATTACAGTGTAGTAGAATTTTTA  1702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1675 GTTGGAGGAAT-GGATGTCTTGGGACAATTTTGCA-GATTATGTTGTTTGCTTAAAATGTAA   1732
Qy  1703 TGTGACTGACATGGAAGCCTGGATGACTATCGTGTTACTTGGGAAAGTCT---CTTTGCTCT  1759
              |||||||||||||||||||||||||||||||||||||||||||||||||   |||||||
Db  1733 TGTAACTGACACAGAGCCCAGAAAGCTAATGTGTTTTTTGGAAAAGTCTCTGCTGAACTGA  1792
Qy  1760 ATTTGCTGACATGCTTCCTGTTGTTGGTCTGGTTGGGCCAAATGCCAAATGTACTCGAATGATGTT-  1818
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1793 ACTTGCTAACATGCTTCCTGTTGTTGTGGGCTGGGCCAGTGTGGTGGTAAATGTACTCCAGTGATGTTA  1852
Qy  1819 AAGGGCTCTGTAAAAACTTCATATACCCTCTTTGGCCATTGTATGCATGATGATTGTTTGGTTTTA  1878
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1853 AAGGGCTCGTGAGAACTTTATACCTCTCTTGGGATTGTGGAGTTTTATAGCCATGCAGTTTGTTTTTCA  1912
Qy  1879 AACATGGGTATAAATGAAATTGTGTACTTCTCTCAGAAGAGAAAGCAGGTAGAAAAGTACTCCAAT  1938
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1913 GATGTGGT---ATGAACTGAGTGGCTTCTGACAGCAAAAAGGCAGGAGGTACTTACTAGAAGGTGTCCAAT  1969
Qy  1939 TAAAAAA  1945
              |||||
Db  1970 TTTTAAA  1976
```

RESULT 2

AK044835

LOCUS   AK044835   3286 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130007B24 product:pre B-cell leukemia
            transcription factor 1, full insert sequence.
ACCESSION   AK044835
VERSION   AK044835.1 GI:26336856
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10349636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3286)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,

```
          Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
          Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
          Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Tagami,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
          Takeda,Y., Tanaka,T., Tonaru,A., Toya,T., Yasunishi,A.,
          Muramatsu,M. and Hayashizaki,Y.
  TITLE
  JOURNAL  Direct Submission
          Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)
  COMMENT  cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in Riken
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Tissues were provided by Dr. Tomohiro Kono (Department of Animal
          Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
          Kanagawa Prefecture, Japan) whose assistance we gratefully
          acknowledge.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
  FEATURES             Location/Qualifiers
       source          1..3286
                       /organism="Mus musculus"
                       /mol_type="mRNA"
                       /strain="C57BL/6J"
                       /db_xref="MGI:2410190"
                       /db_xref="FANTOM_DB:B130007E24"
                       /db_xref="taxon:10090"
                       /clone="B130007E24"
                       /clone_lib="RIKEN full-length enriched mouse cDNA library"
                       /dev_stage="9.5 days embryo"
                       /tissue_type="parthenogenote"
       CDS             123..1685
                       /note="unnamed protein product; pre B-cell leukemia
                       transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
                       evidence: BLASTN, 99%, match=124)"
                       /codon_start=1
                       /protein_id="BAC32112.1"
                       /db_xref="GI:26336857"
                       /translation="MPGSLPLNAEACWPKDVGIVALEIYPPSQYVDQAELEKVDGVDA
                       GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSVDCIGRLEVGTETIIDKSK
                       SVKSNLMQLPEESGNTDIEGIDTTNACYGGTAAVFNVANVWVESSSWDGRYALVVAGI
                       AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHIMQHAYDYFKGPMKLSEXYPVDG
                       KLSIQCYLSALDRCVSYRKIRAQWQKEGKGKDCFTLNDFGFMIFHSPYCKLVQKSLA
                       RMFLNDFLNDDNRDKNSIYSGLRAFGDVKLEDTYFDRDVEKAFMKASSELFNQKTKAS
                       LIVSNQNGNMYTPSSVYGSLASVLAQYSFQQLLAGKRVGVFSYGSGLAATLYSLKVTQDA
                       TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                       GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIFSPAKKVPRLPATSA
                       ESESAVLSNGEH"
       polyA_signal    3272..3277
                       /note="putative"
       polyA_site      3286
                       /note="putative"
  ORIGIN
  Query Match     59.4%;  Score 1189.6;  DB 11;  Length 3286;
  Best Local Similarity   79.9%;
  Matches 1552;  Conservative    0;  Mismatches  244;  Indels  146;  Gaps    7;

Qy        254 AAAAATATGCATGGTGTAGAGTGCTGGGAAAGTATACCATTGGCTTGGGGCCAGGCCAAGGA 313
                 |||| ||||||||||||||| |||||||||  ||||||| ||||||||||| ||||||| ||
Db        232 AAAAATACGAATGGTTGTGATGCCTGGGAAAGTATACCAATCGGCGCCTGGGGCCAGGGGATGG 291

Qy        314 GCTTCTCGCACAGATAGAGAAGAGATTAACTCTCTTTGCATGACTGACTGTGGTTCAGAATCTTA 373
              ||||| || || |||||| |||||||  |||||| |||||| |||||| |||||||||||||||
Db        292 GCTTCTCGCACGGATCGTGAAGACATCAACTCTCTTTGCCTGACTGTGGTTCAGAAGCTAA 351

Qy        374 TGGAGAGAAAATAAACTTTCCTATGAATTGCATTGGCGCCTGGAACAGAACAA 433
              ||| |||||| || | ||||||||||| ||||| ||||| ||||||||||||
Db        352 TGGAGAGACATAGCCTTTCCTATGATTGCATTGGCGGGCCTAGAAGTTGGAACAGAGACA 411

Qy        434 TCATCGACAAATCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGA 493
              |||| |||||||||||||||| ||| ||||| ||||||||||||||| ||||||||||||
Db        412 TCATCGACAAAATCGAAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCTGGGA 471

Qy        494 ATACAGATATAGAAGGAATCGACACTAATGCATCTATGGAGGGCACAGCTGCTGTCTCT 553
              ||||||||| | |||||||| |||| ||||||||| |||||||||||||||| |||||
Db        472 ATACAGATATAGAAGGAATAGATACAACCAATGCATGCTATGGGGGCACAGCTGCAGTCT 531

Qy        554 TCAATGCTGTTTAACTGGATTGAGTCCAGCTCTTGGGAT 591
              ||| ||||||||| ||||||||| |||||||||||||||
Db        532 TCAATGCCGTGAACTGGGTCGGGATGGGACGGATAGTCTCTGGGAGTTG 591

Qy        592 CAGGAGACACATTGCTATATAATGCCACAGGAAAATGCCAGACCTACAGGTGGGAGTGGAGCTG 651
                                 ||||| |||||||||| |||||||||
Db        592 -------------------GGGCTTCGTTGGGACAC 607

Qy        652 TGGGCCCTGCTAATTGGGGCCAAAACGCTCCTCTCTAATTTTTGACCGAGGGGTCCGTGGAGACAC 711
              ||||||||||||||| |||||||||||||||||||||||| |||||||||| ||||| |||
Db        608 ATATCCAACATGCCTATGATTTTACAAGCCTGATATGCTACTACTATAGTAG 667

Qy          ... (additional alignment rows continue) ...
```

```
Qy    1208 CTCTTGATAAAATAAACAGCAAGTTTATGTGGATCTTAAATCAAGGCTTGATTCAAGAACTG 1267
                ||||||||||| ||||||||||||||||||||||||||||||| |||||||||||||
Db    1312 CCCTTGATAAAATAAACAGCAAGTTTATGTGGATCTTAAATCAAGGCTTGATTCAAGAACTT 1371

Qy    1268 GTGTGGCACCGGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 1327
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1372 GTGTGGCACCGGATGTCTTTGCTGAAAACATGAAGCTCAGAGAGGACACCCATCACTTAG 1431

Qy    1328 TCAACTATATTCCCAGGGGTTCAATAGATTCACTCCTTTGAAGGAACGTGGTGTACTTAGTTA 1387
                ||||||||||||||||||||| ||||||||||||||||||||| || |||||
Db    1432 CCAACTATATTCCCAGTGTTCAATAGACTCACTCTTTGAAGGAACGTGGTATCTGGGTCA 1491

Qy    1388 GGGTGGGATGAAGAAAGACACAGAAGAGGACTTACGGTCGTCCCACTCCAAATGATGACACTT 1447
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1492 GAGTGGATGAAAAAACACAGAAGGACTTACGGTCCCGGGCCCTTCACAAATGACCACAGTT 1551

Qy    1448 TGGATGAAGGAGTAGGAACTTGTGCCATTCAAACATAGCAACATGACCATAATTCCAAGCCCTG 1507
                |||||||||||||||||||||||||||||||||| ||||||| |||||||||||||
Db    1492 TGGATGAAGGAATGGGGGTTCGTGCATAGTAACACAGCAACAGGCATAATTCCAAGCCCTG 1611

Qy    1508 CCAAGAAAAGTACCAAGACTCCCTGCCACAGCAGCAAGAACCTGAGCAGTGTGTCATTAGTTA 1567
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1612 CTAAGAAAGTGCCAAGACTCCCTGCCGGAATCGGAATCAGCTGTGTGTCATCAGTA 1671

Qy    1568 ACGGGGGAACATTAAGATACTCTCTGTGAGGGTGCCAGACTTCAGGGTGTGGGGGTGCATGGGGT 1627
                |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1672 ACGGGGAGCACTGAGAGTCTGTGCCCTTTACA-----------GAGGCTCGGGGGCCGGGA 1720

Qy    1628 GGGGGTATGGGAACAGTTGGGAGGAATGGGGATAATATTAAAGGATTAACATG 1687
                ||||||||||||||||||||||||||||||||||||||| ||| || |||||
Db    1721 TGGGGTATGGGAAAACGGTTGGGAGGAAT-GGATTGCTTGGGGACACAATTTTGCA--GAATTATGTTG 1778

Qy    1688 TTATGTAAATTTTATTGTGTGACTGGATGGGAGGCCTGGATGATCGTGTACTTGGGAAAG 1747
                ||| |||| || || |||||||||||||||||||||||| ||||||||||| ||||||
Db    1779 TTGCTTAAAATGTAATGTAGTAACTGACACAGAGCCCAGAAAAGCTATTGGTGTTTTTGGAAAAG 1838

Qy    1748 TCT---CTTTGCTCTATTTGCTCTGTCGTTCCTGTTGTGGTCTGGGTCTGGCCAAATGCCAAAATGT 1804
                |||| || |||||||||||||||| || |||
Db    1839 TCTCTGCTGAACTGAACATGCTTCCTGTTCTGGCTCTGGCCAGTGGTGGTAAATGT 1898

Qy    1805 ACTCGGAATGATGTT-AAGGGCTCTGTAAAACTTCATACTCATTTTGGTTATGCAT 1863
                |||||||||||||| ||||||||| |||||||||||||||| |||||||||||
Db    1899 ACTGCAGTGATGTTAGTTATACCTCGGATGTTTTATACGCAT 1958

Qy    1864 GATGTTTGGTTTTTAAACATGGTATATAATGAATTGTGTACTTCTGTCAGAGAAAGCAGAG 1923
                ||| || |||||||||| ||| || ||| |||| ||| |||||||||||||
Db    1959 GCAGTTTAGTTTCAGATGTGGT---ATGAACTGAGTGCTTCTGACAGCAGAAGGGCAGAG 2015

Qy    1924 GTACTAAATCTCCAATTAAAAA 1945
                |||||||||||||||||||||
Db    2016 GTACTAGTGTCCAATTTTTTAA 2037
```

RESULT 3

```
CDS             122..1684
                /note="unnamed protein product; pre B-cell leukemia
                transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
                evidence: BLASTN, 99%, match=124)
                /putative"
                /codon_start=1
                /protein_id="BAC27338.1"
                /db_xref="GI:26327189"
                /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                GKYTIGLGQARMGFCTDREDINSILCLTVVQAMERHSLSVDCIGRLEVGTETIIDKSK
                SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVNVAVNWVESSSWDGRYALVVAGDI
                AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAPMKKASSELFNQKTKAS
                LIVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRVGVFSYGSGLAATLYSLKVTQDA
                TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                GTWYLVRVDEKHRRTYARPFTNDHSIDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                ESESAVISNGEH"
polyA_signal    3447..3452
                /note="putative"
polyA_site      3466
                /note="putative"
ORIGIN

Query Match        59.4%;  Score 1189.6;  DB 11;  Length 3466;
Best Local Similarity   79.9%;  Pred. No. 1.8e-174;
Matches 1552; Conservative   0; Mismatches 244; Indels 146; Gaps   7;

Qy   134 GCTCTTTCACCATGCCTGGGATCACTCACTTCCTTTGAATGCAGAAGCTTGCTTGGCCAAAGATG 193
         |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   111 GCTCTTTCACCATGCCTGGGTCACTCACTTCCTTTGAATGCAGAAGCTTGCTTGGCCAAAGATG 170

Qy   194 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 253
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 TGGGAATTGTTGCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAGCAGAGTTGG 230

Qy   254 AAAAATATGATGGTGTAGATGCTGGAAAGTATACACCATTGGGTCTTGGGCCCAAGATGG 313
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 AAAAATACGATGGGTGTAGAGTGCTGGAAAGTATACCATCGGCCTGGGCCCAGGAAACTGA 290

Qy   314 GCTTCTCGCACGATAGAGAAGAGATATTAACTCTCTTTGCCATGACTGTGGTTCAGAATCTTA 373
         |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   291 GCTTCTCGCACGGATCGTGAAGAGCATCAACTCTCTTTGCCCTGACTGTGGTTCAGAAACTGA 350

Qy   374 TGGAGAGAGAAATAAACCTTTCCTATGAATTGCATCCATTGGGGGCGTGGAAGTTGGAACAGAGACA 433
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 TGGAGAGAGAAATAAACCTTTCCTATGATTGCCATTGGGGGCGGGCTTAGGAAGTTGGAACAGAGACA 410

Qy   434 TCATCGACAAAATCAAAGTCTGTGAAGACTAATTGATGAGCTGTTGAAGAGTCTGGGA 493
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 TCATCGACAAAATCGAAATCAGTGAAGTCTAATTGATGCAGCTGTTTGAAGAGTCTGGGA 470

Qy   494 ATACACAGTATAGAACGGAATCGACACAACTAATGCATGCTAATGGAGGCACAGCTGTGTCT 553
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 ATACAGATTATAGAAGGAATAGATACAACCAAGCCATGGGGCCGCTATGGGGGACACAGCTGTCT 530

Qy   554 TCAATGCCTGTTTAACTGGGATTGAGTCCAGCTCTTGGGAT----------------- 591
         ||||||||||||||||||||||||||||||||||||||||||||
Db   531 TCAATGCCTGTGAACTGGGGTGGCTGGAAATCCAGCCTCTTGGGATGGACGGATATGCTCTGGTAGTTG 590

Qy   592 ----------------- 591
                                     |||||||||||||||||||||||||||||
Db   591 CAGGAGACAATTGCTATATATGCCACAGGAAATGCCAGACCTACAGGTGGAGTTGGAGCTG 650

Qy   592 ----------------- 591
                                     |||||||||||||||||||||||||||||
Db   651 TGGCCCTGCTGTAATTGGGGCCAAACGCTCCTCTAATTTTTGACCGGAGGGCTCCGTGGGGACAC 710

Qy   608 ATATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTA 667
         |||||||||||||||||||||||||||||||||||||||||
Db   711 ACACTGCAGCATGCCTATGACTTTTACAAGCCTGACATGCTTCTCCGAGTACCCTGTGGTCG 770

Qy   668 ATGGAAAACTCCATACAGTGCTACCTCAGTGCATTAGACCGGCTGCTACTCTGTCTACT 727
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```
```
Db   771 ACGGAAAAGCTCTCCATACAGTGCTACCTCACGGCCCCTTGGACCGGCCGCTCTATTCTGTCTACC 830

Qy   728 GCAAAAAGATCCATGCCCAGTGGCAGAAAGAGGAGGAAATGAATAAAAGATTTTACCTTGAATG 787
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 GCAAAAAGAATGCCCAGTGCCCAGTGGCAGAAAGAGGAGGAAATGAATAAAAGATTTTACCTTGAATG 890

Qy   788 ATTTTTGGCTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 847
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   891 ATTTTTGGCTTTCATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTC 950

Qy   848 GGATGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATATAAAATAGTGATTCTATAGTG 907
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   951 GGATGTTCCTGAATGACTTTCTTAATGATCAAAACAGAGATAAAAACAGTATTATTACAGTG 1010

Qy   908 GCCTGGAAGCCCTTTGGGAGTTAAATTAGAAGAGCACCCTACTTTGAATGGAGATGTGGAGA 967
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1011 GACTGGAAGCCCTTGGGACGTTAAATTAGAAGAGTACTTTGATAGGGATGTGGAAA 1070

Qy   968 AGGCATTTATGAAGCCTAGCTCTGAACTCTTCAGTCAGAAGAAAAACAAAAGGGCATCTTTACTTG 1027
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1071 AGGCATTTATGAAGCCTAGTTCTGAACTCTTCAGTATATGGTTCCCTTGCATCTGTTC 1130

Qy  1028 TATCAAAATCAAAAATAGTTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTC 1087
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1131 TGTCTAATCAGAATGGAAATATGTACAGCACATCCTCTATGGTGTCGGCTTCTGTGTCC 1190

Qy  1088 TAGCACAGTACTCACCTCAGCAAATTAGCACGGGAAGAGAATTGCAAGTTTCTTATGGGTT 1147
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1191 TGGCACAGTACTCACCTCAGCCAGTTGGCAGGGAAGAGAGATTGGAGTGTTTCTTACGGTT 1250

Qy  1148 CTGGGTTGGGCTGCCACTCGTACTCTCCTTAAAGTCACAACAAGAGATGCTACACCGGGGTCTG 1207
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1251 CTGGCTGGGCTGCCCACCGTCTGTACTCCCTTAAAGTCACAACAAGAGACGCCACACAGGATCTG 1310

Qy  1208 CTCTTGATAAATAAACAGCAAGTTTAGTGGATCTTAAATCAAGGCTTGATTCTCAAGAACTG 1267
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1311 CCCTTGATAAAATAAACAGCAAGTTTTATGTGACCTTAAATCAAGGCTTGACTCGGAACTT 1370

Qy  1268 GTGTGGGCACCAGATGTCTTCCGCTGAAAAACATGAAGCTCAGAGGAGGACACCCATCATTCGG 1327
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1371 GTGTGGCCACCGGAGTGTCTTTGCTGAAAAACATGAAGCTCAGAGGAGGACACACTCACTTAG 1430

Qy  1328 TCAACTATATATTCCCCAGGGTTCAATAGATTCACTCTTCTTTGAAGGGAAACGTTACTTAGTTA 1387
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1431 CCAACTATATATTCCCCAGTGTTCAATAGACTCACTCTTCTTTGAAGGAAACGTGGTATCTGGTCA 1490

Qy  1388 GGGTGTGAGTGAAAAGCACAGAGAGAACACTTACGCTCGGGGGTCCCACTCCAAATGATGACTT 1447
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1491 GAGTGGGATGAAAAACACAGAAAGGACTTAGCCCCGCGCCCTTCACAACTGACACAGTT 1550

Qy  1448 TGGATGAAGGACTAGGACTTGTGCATTCAAACATAGCAACTGAGCATAATTCCAAGCCCTG 1507
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1551 TGGATGAAGGAATGGGGCCTCCGTGCATAGTAACACCGGCAACAGCAACACAGTT 1610

Qy  1508 CCAAGAAAGTACCAACTCCCTCCCACGAGCAGCAAAACCTGAAGCCAGTGTTGTCATTAGTA 1567
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1611 CTAAGAAAGTTGCCCAAGAACTCCCTGCCTCGGCCACGTAGCCAGTGTGTGTCATCAGTA 1670

Qy  1568 ATGGGGGACATTAAGATACTCTGTTGAGGTGCAAGACTTCAGGGTGGGGTTGGGCATGGGGT 1627
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1671 ACGGGGAGCACTGAGAGTCTTGTGGGCCTTTACA------------GAGGCTCGGGGGCCGGGA 1719

Qy  1628 GGGGGTTATGGGAACAGTTGGAGGAGATATCTGGGGGATAATTTAAAGGAGATTACATG 1687
         |||||||||||||||||||||||||||||||||||||||||||||
Db  1720 TGGGGGTTATGGAGAAACGGTTGGAGGAAT--GGATGCTCTTCCGTGGGACAC----GATTATGTG 1777

Qy  1688 TTATGTAAATTTTTATGTGACTGGACATGGAGCCTGGATGACTAGCGTGTTGTACTTGGGAAAG 1747
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1778 TTGCTTAAAATGTAATGTAACTGACACAGAGCCCAGGAAGCTATTGTTGTTTTTGGGAAAG 1837

Qy  1748 TCT---CTTTGCTCTTATTTGCTGACATGCTTCCTGTTGTGTTGGGTCTTGGCCAAATGT 1804
         |||    |||||||
```

```
Db   1838 TCTCTGCTGAACTTGCTAACATCGTTTCCTGTTGTGGTGCTCTGGCCAGTGGGTAAATGT 1897
Qy   1805 ACTCGAATGAATGTT-AAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTTGTTATGCAT 1863
Db   1898 ACTGCAGTGATGTAGTTAAAGGGCTCTGTAGAACTTTATACCTCTCCGGATGTTTATACCGGAT 1957
Qy   1864 GATGTTTGGTTTTTTAAACATGGTATATAAGAATTGTGTTACTTCGTTCAGAAGAAAGCACAG 1923
Db   1958 GCAGTTTAGTTTTTCAGATGTGGTT---ATGAACTGAGTGCTTCTGACAGCAGAAAGGCAGAG 2014
Qy   1924 GTACTAATCTCCAATTAAAAAA 1945
Db   2015 GTACTAGTGTCCAAATTTTTAA 2036
```

RESULT 4

```
LOCUS       BQ049460                  960 bp    mRNA    linear     EST 29-MAR-2002
DEFINITION  AGENCOURT_7064727 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5787318
            5', mRNA sequence.
ACCESSION   BQ049460
VERSION     BQ049460.1  GI:19808800
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12877  row: h  column: 07
            High quality sequence stop: 595.
FEATURES             Location/Qualifiers
     source          1..960
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5787318"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH MGC 85"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN
```

Query Match 41.4%; Score 829.4; DB 12; Length 960;
Best Local Similarity 94.2%; Pred. No. 1.2e-118;
Matches 905; Conservative 0; Mismatches 51; Indels 5; Gaps 4;

```
Qy    865 TTCCTTAATGACCAGAATAGAGATAAAAAATAGTAATCTATAGTGGCCTGCCTGCTTTGGG 924
Db      1 TTCCTTAATGACCAGAATAGAGATAAAAATAGTAATCTATAGTGGCCTGCGAAGCCTTTGGG 60
Qy    925 GATGTTAAATTAGAAGAACACTTACTTTGTATGAATGTGGGAAGAGGCATTATGTGAAGGCT 984
Db     61 GATGTTAAATTAGAAGAACACCTACTTTGTATGAGGATGTGGAGAGGCATTATGTGAAGGCT 120
Qy    985 AGCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTACTTGTATCAAATCAAAAATGGA 1044
Db    121 AGCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCACTTACTTGTTATCAAATCAAAAATGGA 180
```

```
Qy   1045 AATATGTACACATCTTCAGTATATGGTTCCGTTGCATTCTGTTCTAGCACAGTACTCACT 1104
Db    181 AATATGTACACATCTTCAGTATATGGTTCCGTTGCATTCTGTTCTAGCACAGTACTCACT 240
Qy   1105 CAGCAATTAGCAGAGGAAGAGAGAATTGGAGTGAGTTCCTTTCTTATTGTTTCTTAATTGGGCTTAT 1164
Db    241 CAGCAATTAGCAGAGGAAGAGAGAATTGGAGTGTTTCTTATGGTTGCTTTGGGCTGGCCACT 300
Qy   1165 CTGTACTTCCTTAAGTCACAACAGATGCTACACGGGGGTGTCGTCTCTTGATAAAAATAACA 1224
Db    301 CTGTACTTCCTTAAAGTCACAACAAGATGCTACACCGGGGTCTGCTTCTTGATAAAAATAACA 360
Qy   1225 GCAAGTTTATTGTGTGTAAAATCAAAGGCTTGATTCAAGAACTGGTTGGGCACCCAGATGTC 1284
Db    361 GCAAGTTTATTGTGATCTTAAAATCAAAGGCTTGATTCAAGAACTGGTTGGGCACCCAGATGTC 420
```

RESULT 5

```
LOCUS       BG697489                  881 bp    mRNA    linear     EST 07-MAY-2001
DEFINITION  602660682F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4803788 5',
            mRNA sequence.
ACCESSION   BG697489
VERSION     BG697489.1  GI:13963764
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
```

```
Qy   1285 TTCGCTGAAAACATGAAGCTTCAGAGGAGGACCACCCATCATTGGTCAACATTATTCCCCAG 1344
Db    421 TTCGCTGAAAACATGAAGCTTCAGAGGAGGAGGACCACCCATCATTGGTCAACATTATTCCCCAG 480
Qy   1345 GGTTCAATGAATTCACTCTTTGAACGGAACCTGGTACTTACTTACTTAGGGTGGATGAAAAVGCAC 1404
Db    481 GGTTCAATAGAATTCACTCTTTGAACGGAACCTGGTACTTAGTTAGGGTGGATGAAAAGCAC 540
Qy   1405 AGAAGAAACTTACGCTCGGGCGTCCCACTCCAAAATGATTGACA-CTTTGGATGAAGGAGTAGG 1463
Db    541 AGAAGAAACTTACGCTCGGGCGTCCCACTCCAAAATGATGACACCTTTGGATGAAGGAGTAGG 600
Qy   1464 ACTTGTGCATTCAAACATAGCAACTAGCAACTGAGCAGTATATTCCAAGCCCTGCCAAGAAGTACCAAG 1523
Db    601 ACTTGTGCATTCAAACATAGCAACTAGCAACTGAGCAGTATATTCCAAGCCCTGCCAAGAAGAAGTACCAAG 660
Qy   1524 ACTCCCTGCCACAGCAGCGAAACCTGAAGCAGCGTCTCATTAGTAATGGAGGAACATTAAGA 1583
Db    661 ACTCCCTGCCACAGCAGCGAAACCAGTTGTCATTAGTAATGGAGGAACATTAAGA 720
Qy   1584 TACT-CTGTGCAGGTGCCAAGACTTCAGGGTGGGTGGGC--ATGGGGTGGGGGGTGGGTAGGGA 1640
Db    721 TACTCCCTGGGAGGTGCCAAGACTTCCGGGTGGGGCGGGCCCATGGGGGGGGGTATGGGGA 780
Qy   1641 CAGTTGGAGGAATGGGATATCTGGGGGATGAATTTAAAGGGATTACCGGGTTAATGTAAATTT 1700
Db    781 CAGTTGGAGGAATGGGATATCT-GGGAAAAATTTTTAAAGGATTACCGGGTAGGGAAATTTT 839
Qy   1701 TATGTGACTGACATGGAGCCTGGATGACTAGTTGACTTGGGGAAAGTCTCTTTTGCTCTTA 1760
Db    840 TATGTGACTGACAGGGGAGCCCGGATGACTAACCGGAAAGCCCCTTGCCCCA 899
Qy   1761 TTTGGCTGACATGCTTCCCTGTTTGTTGGTGTCTGGCCAATGCCAAAATGTACTCGAATGATCGTTAA 1820
Db    900 TTTGCCGGACGGGCTTCCCGGGGGGGGCTGGCCCAAGGGTAATCCAAAAGGAGGGA 959
Qy   1821 G 1821
Db    960 G 960
```

JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10700   row: c   column: 21
            High quality sequence stop: 815.

FEATURES   Location/Qualifiers
     source   1..881
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4803788"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

ORIGIN
Query Match   40.0%;   Score 800.6; DB 12;   Length 881;
Best Local Similarity   97.4%;   Pred. No. 3.5e-114;
Matches   856; Conservative   0; Mismatches 19; Indels   4; Gaps   4;

Qy   964 GAGAAGGCATTTATGAAGGCTAGCCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTA 1023
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GAGAAGGCATTTATGAAGGCTAGCCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTA 60

Qy   1024 CTTGTATCAAAATCAAAAATGGGAAATAATGTACACATCTTTCAGTATATAGTGTTCCCTTGCATCT 1083
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTTGTATCAAAATCAAAAATGGGAAATAATGTACACATCTTTCAGTATATAGTGTTCCCTTGCATCT 120

Qy   1084 GTTCTAGCACAGTACTCACCTCAGCAATTACCACGGGAAGAGAGAAATTGGGAGTGTGTTTTCTTAT 1143
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTTCTAGCACAGTACTCACCTCAGCAATTACCACGGGAAGAGAATTGGGAGTGTGTTTTCTTAT 180

Qy   1144 GGGTTCTGGTTTGGCTGCCCACTCTGTACTCTCTCTTAAAGTCACACAAGAGCTACACGGGGG 1203
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGGTTCTGGTTTGGCTGCCCACTCTGTACTCTCTTAAAGTCACACAAGAGCTACACGGGGG 240

Qy   1204 TCTGCTCTTGATAAAAATAACACAGCAAGTTTATGTGAGTCAGAAGTTTAATTATTTGATTCAAGA 1263
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TCTGCTCTTGATAAAAATAACACAGCAAGTTTATGTGAGTCAGAAGTTTAATTATTTGATTCAAGA 300

Qy   1264 ACTGGTGTGGCACCAGAGTGTCTTCGCTGAAAAACATGAAGCTCACAGAGGACACCCATCAT 1323
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACTGGTGTGGCACCAGAGTGTCTTCGCTGAAAAACATGAAGCTCACAGAGGACACCCATCAT 360

Qy   1324 TTGGTCAACTATATTCCCCAGGGTTCAATAGAAGTCACTCTTTGAAGGAACGTGGTACTTA 1383
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTA 420

Qy   1384 GTTAGGGTGGATGAAAAGCACAGAAGAACTTACGCTTCGGGGGTCCCACTCCAAATGAGTGAC 1443
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTTAGGGTGGATGAAAAGCACAGAAGAACTTACGCTTCGGGGGTCCCACTCCAAATGATGAC 480

Qy   1444 ACTTTGGATGAAGCAGAGTAGGACTTGTGCCATTC-AAACATTAGCAACTGACCATATTCCAAG 1502
            ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   481 ACTTTGGATGAAGGAGTAGGACTTGTGCCATTCAAAAACATTAGCAACTGACCATATTCCAAG 540

Qy   1503 CCCTGCCACAGAAAGTACCANAGACTCCCTGCCACACGACAGCAGAAACCTGAAGCACCTGTCAT 1562
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CCCTGCCACAGAAAGTACCAAGACTCCCTGCCACACGACAGCAGAAACCTGAAGCAGCTGTCAT 600

Qy   1563 TAGTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGTGGGGTGGGCAT 1622
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TAGTAATGGGGAACATTAAGATACTCTGTGAGGGTGCAAGACTTCAGGGTGTGGGGTGGGCAT 660

RESULT 6
LOCUS   BU162936   1053 bp   mRNA   linear   EST 04-SEP-2002
DEFINITION AGENCOURT_6911057 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:5952255
            5', mRNA sequence.
ACCESSION   BU162936
VERSION   BU162936.1   GI:22676846
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1053)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2139   row: p   column: 16
            High quality sequence start: 3
            High quality sequence stop: 693.

FEATURES   Location/Qualifiers
     source   1..1053
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5952255"
                     /tissue_type="ductal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_110"
                     /note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming;
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

ORIGIN
Query Match   39.4%;   Score 788.6; DB 13;   Length 1053;
Best Local Similarity   94.5%;   Pred. No. 2.3e-112;
Matches   861; Conservative   0; Mismatches 44; Indels   6; Gaps   4;

Qy   896 GTATCTATAGTGGCCTGGAAGCCCTTTGGGGGATGTTAAATTAGAAGAGCAACCTACTTTTGATA 955
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5 GTGGCAATTCCGACGAGGAAGCCCTTTGGGGGATGTTAAATTAGAAGAGCAACCTACTTTTGATA 64

Qy   1623 GGGGTGGGGGTATGGGGAACACAGTTGGAGGAAT-GGGATATCTGGGGATAATTTAAAGGAT 1681
            |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   661 GGGGTGGGGGGTATGGGGAACACAGTTGGAGGAATGGGGGATATCTGGGGATAATTTAAAGGAT 720

Qy   1682 TA-CATGTTATGTAAAATTTTATGTGTGACTGACACATGGCAGCCTGGATGACGTATCGGTGTACTT 1740
            || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TACCATGTTATGTAAAATTTTATGTGTGACTGACACATGGCAGCCTGGATGACGTATCGGTGTACTT 780

Qy   1741 GGGAAAAGTCTCTTTGCTCTATTTTGCTGACACATGCCTTCCGTTGGTGTGGCCAAATGCCAA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GGGAAAGTCTCTTTGCTCTATTTTGCTGACATGCCTTCCGTTGGTGTGGCCAAATGCCAA 839

Qy   1801 ATGTGACTCGAATGAATGTTAAGGGCCTCTGTAAAACTTCAT 1839
            ||||||||||||||||||||||||||||||||||||||||||
Db   840 ATGTGACTCGCAAGGCTGTTAAGGGTGCTGGAAAAATTTAT 878

```
RESULT 7
LOCUS       BG740730                 817 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602615591F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776598 5',
            mRNA sequence.
ACCESSION   BG740730
VERSION     BG740730.1  GI:14051383
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 817)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10629  row: f  column: 23
            High quality sequence stop: 794.
FEATURES             Location/Qualifiers
     source          1..817
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4776598"
                     /clone_lib="NCI_CGAP Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN

Query Match     38.3%;  Score 765.8;  DB 12;  Length 817;
Best Local Similarity   98.8%;  Pred. No. 8.5e-109;
Matches  803;  Conservative    0;  Mismatches   7;  Indels   3;  Gaps   3;
```

```
Qy   956  GAGATGTGGAGAAGGCATTTATGAAGGCTAGTCTTCTGAACTCTTCAGTGTCAGTAAAACAAAGG  1015
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    65  GAGATGTGGAGAAGGCATTTATGAAGGCTAGTCTTCTGAACTCTTCAGTGTCAGTAAAACAAAGG   124

Qy  1016  CATCTTTACTTGTATCAAATCAAAATGAAAATGTACACATCTTCAGTGTCAGTTATAATGGTTCCC  1075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125  CATCTTTACTTGTATCAAATCAAAATGAAAATGTACACATCTTCAGTGTCAGTTATAATGGTTCCC   184

Qy  1076  TTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAGAATTGGAGTGT  1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185  TTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGAGAGAATTGGAGTGT   244

Qy  1136  TTTCTTATGGTTCTCGGTTTGGGGTGCGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTA  1195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245  TTTCTTATGGTTCTCGGTTTGGGGTGCGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTA   304

Qy  1196  CACCGGGGGTCGCTGCTCTTGGTTTGGGCTGCCACTCTGTACTCTCTTAAATCAAGGCTTG  1255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305  CACCGGGGGTCTGCTCTTGGTTTGGGCTGCCACTCTGTACTCTCTTAAATCAAGGCTTG   364

Qy  1256  ATTCAAGAAACTGGTGTGTGGCACCAGATGTCTTCGCTGCTGAAAACATGAAGCTTCAGGACCA  1315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365  ATTCAAGAAACTGGTGTGTGGCACCAGATGTCTTCGCTGCTGAAAACATGAAGCTTCAGGACCA   424

Qy  1316  CCCATCAATTTGGTCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGT  1375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425  CCCATCAATTTGGTCAACTATATTCCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGT   484

Qy  1376  GGTACTTAGTTAGGGTGGGATGGATGAAAAGCACAGAAGAAGAACTTAGGCTTCCCACTCCA  1435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   485  GGTACTTAGTTAGGGTGGGATGGATGAAAAGCACAGAAGAAGAACTTAGGCTTCCCACTCCA   544

Qy  1436  ATGATGACACTTTGGATGAAGGAGTAGGACTTGTGCCATTCAAATCAAGCAAGCTGAGCATA  1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   545  ATGATGACACTTTGGATGAAGGAGTAGGACTTGTGCCATTCAAATCAAGCAAGCTGAGCATA   604

Qy  1496  TTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCGAGAACCTGAAGCAG  1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   605  TTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCCTGCCACAGCAGCAGCGAGAACCTGAAGCAG   664

Qy  1556  CTGTCATTAGTTAGGGTGGGATGGATGAAAAGCACAGAAGAAGAACTTAGGCTTCCCACTCCA  1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   665  CTGTCATTAGTTAGGGTGGGATGGATGAAAAGCACAGAAGAAGAACTTAGGCTTCCCACTCCA   724

Qy  1616  TGGGCAT-GGGGTGGGGTATGGGGATAATGGGGATATCTGGGGATAATTTT  1674
          ||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   725  TGGGCATGGGGGTGGGGTATGGGGATAATGGGGATATCTGGGGATAATTTT   784

Qy  1675  AAAGGATTTACATGTTAT-GTAAATTTTATGTGGACATGACATGG-AGCCTGGATGACTATC  1732
          |||||||||||||||||| ||||||||||||||||||||||||| |||||||||||||||
Db   785  AAAGGATTTACATGTTATGTAAATTTTATGTGGACATGACATGGAAGCCTGGATGACTATC   844

Qy  1733  GTGTACTTGGGAAAGTCTCTTTGCTCT---ATTTGCTGACATGCCTTCCTGTTGTTGGTCTG  1789
          |||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db   845  CTGTACTTGGGCACAGTCTTCTTTAATTGCTGACATGCCTTCCTGTTGTTGGTCTG   904

Qy  1790  GCCAATGCCAA  1800
          |||||||||||
Db   905  TCTGGGCCCAA   915
```

```
RESULT 8
BG742704
LOCUS       BG742704                738 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602632l9F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778093 5',
            mRNA sequence.
ACCESSION   BG742704
VERSION     BG742704.1  GI:14053357
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10633  row: e  column: 06
            High quality sequence stop: 738.
FEATURES             Location/Qualifiers
     source          1..738
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4778093"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Query Match             36.8%;  Score 737;  DB 12;  Length 738;
Best Local Similarity   100.0%;  Pred. No. 2.5e-104;
Matches  737;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1675 AAAGGATTACATGTTATGTAAATTTTTATGTGACTGAACATCATGGACGCCTGGATGACTATCGT 1734
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       600 AAAGGATTACATGTTATGTAAATTTTTATGTGACTGAACATCATGGACGCCTGGATGACTATCGT 659

Qy      1735 GTACTTGGGAAAAGTCTTCTTTGCTCCTATTTGTGTGACATGCTTCCTTCCTTGTTGGGTCTGGCCAA 1794
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       660 GTACTTGGGAAAAGTCTTCTTTGCTCCTATTTGTGTGACATGCTTCCTTCCTTGTTGGGTCTGGCCAA 719

Qy      1795 TGCC-AAATGTACTTCTGAATGATGTTAAGGGCTTCTGTAAAAC-TTCATACCTCTTTGCCA 1852
             |||| |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       720 TGCCAAAATGTACTCGAATGATGTTAAGGGCTTCTGTAAAACTTCATACCTCTTTGCCA 779

Qy      1853 TTTGTATGCATGATGTTTGGTTTTTAAACATGG 1885
             ||||||||||||||||||||||||||||||||
Db       780 TTTGTATGCATGATGTTTAGGGTTTTAAACCTTGG 812
```

```
RESULT 9
BG683150
LOCUS       BG683150                894 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602651636F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761657 5',
            mRNA sequence.
ACCESSION   BG683150
VERSION     BG683150.1  GI:13914547
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1614  row: h  column: 10
            High quality sequence stop: 799.
FEATURES             Location/Qualifiers
     source          1..894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

Qy      1378 TACTTAGTTAGGGTGGGATGGAAAAAGCACAGAAGAAACTTACGGTCGGTCCCACTCCAAAT 1437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TACTTAGTTAGGGTGGGATGGAAAAAGCACAGAAGAAACTTACGGTCGGTCCCACTCCAAAT 240

Qy      1438 GATGACACTTTGGATGAAGGAGTAGGAGACTTGTTGCATTCAAAACATAGCAACTGAGCATATT 1497
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GATGACACTTTGGATGAAGGAGTAGGAGACTTGTTGCATTCAAAACATAGCAACTGAGCATATT 300

Qy      1498 CCAAGCCCTGCCCAAGAAGAGTACCCAAGACTCCCTGCCCACAGCCAGCGAGCAGAACCTGAAGCAGCT 1557
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CCAAGCCCTGCCCAAGAAAGTACCCAAGACTCCCTGCCCACAGCCAGCGAGCAGAACCTGAAGCAGCT 360

Qy      1558 GTCATTAGTTAAATGGGGAACATTAAGATAATACTTCTGTGGAGGTCGAAGACTTCAGGGGTGGGGGTG 1617
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GTCATTAGTTAAATGGGGAACATTAAGATACTTCTGTGGAGGTCGAAGACTTCAGGGGTGGGGGTG 420

Qy      1618 GGCCATCAGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGGATAATCTGGGGATAATTTTAAA 1677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GGCATCAGGGGTGGGGGTATGGGAACAGTTGGAGGAATGGGGATATCTGGGGGATAATTTTAAA 480

Qy      1678 GGATTACATGTTATGTAAAATTTTTATGTGACTGAACATCATGGACGCCTGGATGACTATCGTGTA 1737
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GGATTACATGTTATGTAAAATTTTTATGTGACTGAACATCATGGACGCCTGGATGACTATCGTGTA 540

Qy      1738 CTTGGGAAAGTCTCTTTGCTCTATTTGCTCGACATGCTTCCTTGTTGGTTGTGGGTCTGGCCAATGC 1797
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CTTGGGAAAGTCTCTTTGCTCTATTTGCTCGACATGCTTCCTTGTTGGTTGTGGGTCTGGCCAATGC 600

Qy      1798 CAAAATGTACTCGAATGATGTTAAGGGGCTCTGTAAAACTTCATACCTCTTTGGCCCATTTGT 1857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CAAAATGTACTCGAATGATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTTGT 660

Qy      1858 ATGCATGATGTTTGGTTTTTAAACATGGTATAAATGTGTACTTCTGTACTTCTGTCAGGAAGAAA 1917
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 ATGCATGATGTTTGGTTTTTAAACATGGTATAAATGTGTACTTCTGTACTTCTGTCAGAAGAAA 720

Qy      1918 GCAGAGGTACTAATCTC 1934
             |||||||||||||||||
Db       721 GCAGAGGTACTAATCTC 737
```

```
RESULT 9
BG683150
LOCUS       BG683150                894 bp    mRNA    linear   EST 01-MAY-2001
DEFINITION  602651636F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761657 5',
            mRNA sequence.
ACCESSION   BG683150
VERSION     BG683150.1  GI:13914547
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1614  row: h  column: 10
            High quality sequence stop: 799.
FEATURES             Location/Qualifiers
     source          1..894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

Qy      1198 CCGGGGGTCTGCTTCTTGGATAAAATAAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 1257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CCGGGGGTCTGCTTCTTGGATAAAATAAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 60

Qy      1258 TCAAGAACTGGGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGAGACACC 1317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 TCAAGAACTGGGTGTGGCACCAGATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGAGACACC 120

Qy      1318 CATCATTTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCCTTTGAAGGAACCTGG 1377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 CATCATTTGGTCAACTATATTCCCCAGGGTTCAATAGATTCACTCCTTTGAAGGAACCTGG 180
```

```
            /db_xref="taxon:9606"
            /clone="IMAGE:4761657"
            /tissue_type="neuroblastoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH_MGC_47"
            /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies).
            Note: this is a NIH_MGC Library."
ORIGIN
```

Query Match   35.3%;   Score 706.2;   DB 12;   Length 894;
Best Local Similarity   92.8%;   Pred. No. 1.3e-99;
Matches 831;   Conservative   0;   Mismatches   48;   Indels   16;   Gaps   8;

```
Qy   1098 CTCACCTCAGCAATTAGCAGGGGAGGGGAGGAAGAAGTTGGAGTTGTGTTTCTTATGGTTCTGGGTTGGC 1157
          ||||||||||| ||||||| ||| ||||||||||||| |||||| || ||||||| |||||||||||||||
Db      2 CTCACCTCAGCAAATTAGCAGAGAGGAGGAAGAGGAGGAACTGGAGTGTTTCTTAATGGTTCTGGGTTGGC   61

Qy   1158 TGCCACTCTGTACTCTCTTAAAGTCACACAAAGATGCTACACACCGGGGTCTGCTCTTGATAA 1217
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     62 TGCCACTCTGTACTCTCTTAAAGTCACACAAAGATGATGCTACACACCGGGGTCTGCTCTTGATAA  121

Qy   1218 AATAACACGCAAGTTTATGTGATCTTAAATCAAGGCCTTGATTCAAGGACTGGTGTGGGCACC 1277
          |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||| |||||
Db    122 AATAACACCAAGTTTATGTGATCTTAAATCAAGGCCTTGATTCAAGAACTGGTGTGGCACC  181

Qy   1278 AGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAACTATAT 1337
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    182 AGATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGCACACCCATCATTTGGTCAACTATAT  241

Qy   1338 TCCCCAGGGTTCAATAGATTCACTCCTTTGAAGGAAGGTGGTACTTAGTTAGGGTTGGATGA 1397
          ||||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    242 TCCCCA-GGTTCAATAGATTCACTCCTTTGAAGGAAGGTGGTACTTAGTTAGGGTTGGATGA  300

Qy   1398 AAAGCACGAAGAAGAACTTACGGCTCGGCGGTCCCCACTCCCAAATGATCACACTTTGGATGAAGG 1457
          ||||||||||||||||||||| ||||| || |||| |||||||||||| |||||||| |||||||
Db    301 AAAGCACGAAGAAGAACTTACGCTCGGGCGTCCCACTCCCAAATGATGACACTTTGATGAAGG  360

Qy   1458 AGTAGGAGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCCAAGAAAGT 1517
          ||||||| |||||| ||||||| |||||||||||||||||||||||||||||||||||||||
Db    361 AGTAGGACTTGTGCCATTCAAAACATAGCAACTGAGCATATTCCAAGCCCTGCCCAAGAAAGT  420

Qy   1518 ACCAAGACTCCCCTGCCACAGCACAGCAGCAGCTGTCATTAGTTAATGGGGGAACA 1577
          |||||||||||| |||| ||||||||||||||||||||||||| ||||||||| ||
Db    421 ACCAAGACTCCCTCCACACAGCAGCAGCAGCTGTCATTAGTTAGTTAATGGGGGACA  480

Qy   1578 TTAAGATACTCTCTGTGAGGTGCAAGACTTCAGGGTGGGGGTGGGGGGTATGG 1637
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTAAGATACTCTCTGTGAGGTGCAAGACTTCAGGGTGGGGGTGGGGGGTATGG  540

Qy   1638 GAACAGTTGGAGGAATGGGATATCTGGGGGGATAATTTTAAAGGATTACATGTTATGTAAAT 1697
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAACAGTTGGAGGAATGGGATATCTGGGGGGATAATTTTAAAGGATTACATGTTATGTAAAT  600

Qy   1698 TTTTATGTGACTGACATGGAGCCTGTTGTACTTGGTACTTGGGAAAGTCTC--TTTGC 1756
          |||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
Db    601 TTTTATGTGACTGACATGGAGCCTGTTGTACTTGGTACTTGGGAAAGTCTC--TTTGC  660

Qy   1757 TCTATTTGTTGCTGACATGCTTCCTGTTGTGGTTGGTCTGGCCAAATGTACTCCGAATGATG 1816
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCTATTTGTTGCTGACATGCTTCCTGTTGTGGTTGGTCTGGCCAAATGTACTCCGAATGATG  719

Qy   1817 TTAAGGGGCTCGTGTAAA--ACTTCATACC--TCTTGGCCATT--GTATGCATGAGTGT--T 1869
          |||||||||||||||||   |||||||||   ||||||||||   ||||| |||||| ||  |
Db    720 TTAAGGGGCTCGTGTAAATACTTCATACCCTTTTGGCCATTTTGTATTGCCTGAGTGTCCT  779

Qy   1870 TGGTTTTTAAACATGGTATAATGTTGTACTTCTGTC-----AGAAGAAAGCAGAG 1923
          |||||||||||||||||||||||| |||||||||||     |||||||||||||
Db   1870 TGGTTTTTAAACATGGTATAATGTTCTTTGGGCTCGTGTAAATACTTCATACCTTTTGTATTGCCTGATGTCCT  1923
```

```
Db    780 GGTCGTTTAAACATGGTATAATGATCATGAACTGAGAATGATCAAGAACTTCGTGTGTACTTTCCTTGGTGTCAGAAGAAAGCAGAG  839
          |||||||||||||||||||||  ||| |||||||||||||| |||||||||||||||||||||| |||||||||||||||||||||||
Qy   1924 GTACTTAATCTCCAATTAAAAAAAATTTTTTAACATGTAGTGTAAAAAAAAAAAAAAAAAAA 1978
          ||||||||||||||||||||                  |||||||||||||||||||||||
Db    840 GTACTTAATCTCCAATTAACAACATTTTCTCACCTGGTTCAACAAAAAAAAAAAAAA  894
```

RESULT 10

AU127534
AU127534

```
LOCUS       AU127534      703 bp     mRNA    linear    EST  01-AUG-2002
DEFINITION  AU127534 NT2RP2 Homo sapiens cDNA clone NT2RP2001505 5', mRNA
            sequence.
ACCESSION   AU127534
VERSION     AU127534.1   GI:10952250
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 703)
AUTHORS     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
TITLE       HRI human cDNA project
JOURNAL     Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project.; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..703
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RP2001505"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RP2"
                     /note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
                     cells after 2-weeks retinoic acid (RA) induction"
ORIGIN
```

Query Match   33.3%;   Score 666.8;   DB 9;   Length 703;
Best Local Similarity   97.0%;   Pred. No. 1.7e-93;
Matches 682;   Conservative   0;   Mismatches   20;   Indels   1;   Gaps   1;

```
Qy    823 AAACTGGTTCAGAAATCTTCAGTCTCGGATCGTGTTGCTGAATGACTTCCTTAATGACCAGAAT 882
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      1 AAACTGGTTCAGAAATCTTCAGTCTCGGATCGTTGCTGAATGACTTCCTTAATGACCAGAAT   60

Qy    883 AGAGATAAAAATAGTATCTATAGTGGCCTGGAAGGCCTTGGGGATGTTAAATAGAAGAC 942
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     61 AGAGATAAAAATAGTATCTATAGTGGCCTGGAAGGCCTTGGGGANGTTAAATAGAAGAC  120

Qy    943 ACCTACTTTGATAGAGATGTGGAGAAGGCATTATGAGAGGCTAGCTTCTGAACTCTTCAGT 1002
          ||||||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||
Db    121 ACCTACTTTGATAGANATGTGGAGANAGGCATTATGAGAGGCTAGCTTCTGAACTCTTCAGT  180

Qy   1003 CAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAAAAATGGAAAATATGTACACATCTTCA 1062
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAAAAATGGAAAATATGTACACATCTTCA  240

Qy   1063 GTATATGGTTCCCTTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCACGGGAG 1122
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTATATGGTTCCCTTGCATCTGTTCTAGCACAGTACTCACCTCAGCAATTAGCACGGGAAN  300
```

```
Query Match          33.2%;  Score 664.2; DB 13;  Length 1083;
Best Local Similarity 86.2%;  Pred. No. 3.6e-93;
Matches 799;  Conservative  2;  Mismatches  0;  Indels 126;  Gaps  1;

Qy  134 GCTGCTTTCACCATGCCATGGACTCACTTCCTTTGAATGCACGAAGCTTGCTGCCAAAGATG 193
        ||||||| |||||||| ||  :|||||| |||||||||||||||||| ||||||||||| ||||
Db  157 GCTCTTTTCACCATGCCGGGMTCACTTCACTTCCTTTGAATGCACGAGAGCTTGCTGCCAAAAGATG 216

Qy  194 TGGGAAATTGTTGTCCCCTTGAGATCTATTTTTCCTTCTCAATATATGTTGATCAAGGATTGG 253
        ||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||| |||
Db  217 TGGGAAATTGTTGGCCCTTGAGATCTATTTTTCCTTCTCAATATATGTTGATCAAGGAGTTGG 276

Qy  254 AAAAAATATGATGGTGTAGGATGCTGGAAAGTATACCATTGGTGTGGGCCGCCAAGAATGG 313
        ||||||||||||||||||||||||||||||| ||||||||||  ||||||||||||| |||
Db  277 AAAAAATATGATGGTGTAGGATGCTGGAAAAGTATACCATTGGCTGGGGCCGCCAAGATGG 336

Qy  314 GCTTCTGCACAGTAGGAGGAAGAGATTATTAACTCTCTTTTGCATGACTGACTGGTTCAGGAATCTTA 373
        |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db  337 GCTTCTGCACAGTAGGAGGAAGAGATTATTAACTCTCTTTGCATGACTGACTGGTTCAGGAATCTTA 396

Qy  374 TGGGAGAGAAATAACCTTTCCTATGATTGCATTGGCGGCGGCTGGAACAGAGACAA 433
        ||||||||||||||||||||||||||||||||||||||| |||||| ||||||||
Db  397 TGGGAGAGAAATAACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGAGACA 456

Qy  434 TCATCGACAAAATCAAAAGTCTCTGTGAAGACTAATTTGATGCAGCGTGTTTGAAAGTCTGGGA 493
        ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  457 TCATCGACAAAATCAAAAGTCTCGTGAAGACTAATTTGATGCAGCGTGTTTGAAAGTCTGGGA 516

Qy  494 ATACAGATATAGAAAGGAAATCGACACAACTAATGCAATGCTAATGGAGGCACAGCTGCTGTCTCT 553
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  517 ATACAGATATAGAAAGGAAATCGACACAACTAATGCAATGCTAATGGAGGCACAGCTGCTGTCTCT 576

Qy  554 TCAATGCTCGTTAACTGGATTGAGTCCAGCTCTTGGGAT------------------ 591
        |||||||||||||||||||||||||||||||||||||||
Db  577 TCAATGCTCGTTAACTGGATTGAGTCCAGCTCTTGGGATGGACGCTCTTAGCCCTGGGTAGTTG 636

Qy  592 ------------------ 591
Db  637 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTTACAGGTGGGAGTTGGAGCAG 696

Qy  592 CAGGAGATATTGCTGTATATGCCACAGGAAAATGCTAGACCTTACAGGTGGGAGTTGGAGCAG 607
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  697 TAGCTCTGCTTAATTGGGGCCAAATGCTCCTTTAAATTTTGAAACGAGGGCTTCGTGGGGACAC 756

Qy  608 ATATGCAACATGCCTATGATTTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 667
Db  757 ATATGCAACATGCCTATGATTTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAG 816

Qy  668 ATGGAAAACTCTCCATACAGATGCTGACCTTGCATTAGACCGGTTGCTACTCGTCGTCTACT 727
Db  817 ATGGAAAACTCTCCATACAGATGCTGACCTTGCATTAGACCGGTTGCTACTCGTCGTCTACT 876

Qy  728 GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAAATGANTAAAGATTTTACCTTGAATG 787
Db  877 GCAAAAAGATCCATGCCCAGTGGCGAGAAAGAGGGAAAATGATTAAAGATTTTACCTTGAATG 936

Qy  788 ATTTTTGGCTTCATGATCTTTCACTCACCATTATTGTAAACTGGTTCAGGAAATCTCTAGCTC 847
Db  937 ATTTTTGGCTTCATGATCTTTCACTCACCATTATTGTAAACTGGTTCAGGAAATCTCTAGCTC 996

Qy  848 GGATGTTGCTGAATGACACTTCCTTAATGACCAGAATAGAGATAAAATAGTATCTATAGTG 907
Db  997 GGATGTTGCTGAATGACACTTCCTTAATGACCAGAATAGAGATAAAATAGTATCTATAGTG 1056

Qy  908 GCCTGGAAGCCTTTGGGGATGTTAAAT 934
Db 1057 GCCTGGAAGCCTTTGGGGATGTWAAAT 1083
```

```
Qy 1123 AGAATTGGAGTGTTTCTTATGTTGGTTCTGGTTGGCTGCCACTCGTACTCTCTTAAAGTC 1182
Db  301 ANAATTGGAGTGTTTCTTATGGTTCCTGGTTGGCTGCCACTCGTACTCTCTTAAAGTC 360

Qy 1183 ACACAAGATGCTACACCGGGGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTCATCTT 1242
Db  361 ACACAAGATGCTACACCGGGGGTCTGCTCTTGATAAAATAACAGCAAGTTTATGTGATCTT 420

Qy 1243 AAATCAAGGCTTGATTCAAGAAACTGGGTGTGGCACCCAGAGTCTTCGGTCTGAAAACATGAAG 1302
Db  421 AAATCAAGGCTTGATTCAAGAAACTGGGTGTGGCACCCAGAGTCTTCGGTCTGAAAACATGAAG 480

Qy 1303 CTCAGAGAGGAGCACCCATCATTTGGTCAACTATATATTCCCCAGGGTTCAAATAGATTCACTC 1362
Db  481 CTCAGAAANGACACCCATCATTTGGTCAACTATATATTCCCCAGGGTTCAAATAGATTCACTC 540

Qy 1363 TTTGNAGGAGAACCTGGGTACTTAGTTAGGGGGTGGGAGAAAAGCACAGAAGAACTTACCGCTCGG 1422
Db  541 TTTGAAAGAACGGTGGGTACTTANTTAGGGTGGATGAAAAGCACAGAAAAACTTACCGCTCGG 600

Qy 1423 CGTCCCACTCCAAATGAATGATGACACTTTGGATGAAGGAGTAGGACTTGTTGTCATTCAAAACA-T 1481
Db  601 CGTCCCACTCCAAATGATGACACTTTGGATGAAGAGAAATAGGACTTGTTCATTCNAACATT 660

Qy 1482 AGCAAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAGA 1524
Db  661 ANCAACTGAACATATTCCAANCCCTGCCAANAAAGTACCAANA 703
```

RESULT 11
BX441322

```
LOCUS       BX441322          1083 bp    mRNA    linear    EST 15-MAY-2003
DEFINITION  BX441322 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF016YJ04 5-PRIME, mRNA sequence.
ACCESSION   BX441322
VERSION     BX441322.1    GI:30789948
KEYWORDS    EST.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1083)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            This library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3098.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF016DE02QP1&cluster=3098.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fullength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF016DE02QP1.
FEATURES             Location/Qualifiers
     source          1..1083
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF016YJ04"
                     /tissue_type="FETAL BRAIN"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /dev_stage="fetal"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

RESULT 12
BG683791/c
```
LOCUS       BG683791          795 bp    mRNA    linear    EST 01-MAY-2001
```

DEFINITION   602651636T1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761657 3',
              mRNA sequence.
ACCESSION   BG683791
VERSION   BG683791.1 GI:13915188
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 795)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1614 row: h column: 10
              High quality sequence start: 758.
              High quality sequence stop: 24
FEATURES   Location/Qualifiers
     source   1..795
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4761657"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_47"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
                     Note: this is a NIH_MGC Library."

ORIGIN
Query Match   32.8%; Score 657.6; DB 12; Length 795;
Best Local Similarity   95.0%;
Matches 756; Conservative   0; Mismatches 29; Indels 11; Gaps 7;

Qy   1117 GGGAAGAGAAGAATGGAGTGGTGTTTCTTAGGGTTCTCGGTTTGGCTGGCCAACTCTGTACTCTCTT 1176
              |||||||||||||||||||| |||||||| |||||||||||| |||||||||||||||||||||||||
Db   794 GGAACGAGAACCGGAGTGTTTCCT--AGGTTCTGGCTGGCTCCCAACTCTGTACTCTCGGT 737

Qy   1177 AAAGTCACACAAGAGCTACACCGGGGGTGTCGCTCTTGAT-AAAATAACGACAAGTT---T 1232
              |||||||||||||||||||||||||||||||||||||||| || ||||||||||| |
Db   736 AAAGTCACACAGACTGCTACACCGGGGTCGTCGGTCTTGATCCAAATAACGACAAGTTTAT 677

Qy   1233 ATGTGATCTTAAATCAAGGCTTGATTCAA--GAACTGGTGTGGCCACCAGATGTGTTCGCT 1290
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   676 GTGACTCTTAAACTCAAGGCTTGATTCAAAGAACTGGTGTGGCCACCAGATGTCTTCGCT 617

Qy   1291 GAAAACATGAAGCTCAGAGAGGACACCCATCATCATTGGTCAACTATATTCCCCAGGGTTCA 1350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   616 GAACACATGAAGCTCAGAGAGGACACCCATCATCATTGGTCAACTATATTCCCCAGGGTTCA 557

Qy   1351 ATAGATTCACTCTTTGAAGGAACGTGGTAC-TTAGTTAGGGTGGATGAAAAGCACAGGAG 1409
              |||||||||||||||| ||||||||||||| |||||||||||||||||||||||||||||
Db   556 ATAGATTCACTCTTTGAAGGAACGTGGTACTTTAGTTAGGGTGGATGAAAAGCACAGGAG 497

Qy   1410 AACTTACCCTCGCGGCGTTCCCACTCCAAAATGATGACACTTTGGATGAAGGAGTAGGACTTGT 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 AACTGACGCTCGGGCGGTCCCACTCCAAATGATGACACTTTGGATGAAGGAGTAGGACTTGT 437

Qy   1470 GCATTCAAACATAGCAACTGAGCAATATTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCC 1529
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 GCATTCAAACATAGCAACTGAGCAATATTCCAAGCCCTGCCAAGAAAGTACCAAGACTCCC 377

Qy   1530 TGCCACAGCAGCAGAACCTGAAGCAGCAGCTGTCATTAGTAATGGGGAACATTAAGATACTCT 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376 TGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACATTAAGATACTCT 317

Qy   1590 GTGAGGTGCAAGACTTCAGGGTGGGGGTGGGCCATGGGGGGTGTGGGGGGTAGGGAACCAGTTGGAG 1649
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   316 GTGAGGTGCAAGACTTCAGGGTGGGGGTGGGCCATGGGGGGTGTGGGGGGTAGGGAACCAGTTGGAG 257

Qy   1650 GAATGGGGATATCTGGGGAATAATTTTAAAGGATTACATGTTATGTAATTTTTATGTGACT 1709
              ||||||||||||||||||||||||| || |||||||||||||||||||||||| ||||||||
Db   256 GAATGGGGATATCTGGGGGATAATTTTTAAAGGATTACATGTTATGTAATTTTTATGTGACT 197

Qy   1710 GACAATGGGAGGCCTGGATGACTATCCGTACTTGGGAAAGTCTCTTTGCTGTCTATTTGCTGAC 1769
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   196 GACAATGGGAGGCCTGGATGACTATCGTGTACTTGGGAAAGTCTCTTTGCTGTCTATTTGCTGAC 137

Qy   1770 ATGCTTCCTGTTGTGGGTCTGGCCAATGCCAAATGTACTCGGAATGAGTGAAGTTAGGGGC-TCTG 1828
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||||
Db   136 ATGCTTCCTGTTGTGGGTCTGGCCAATGCCAAATGTACTCGGAATGAGTGAAGTTAGGGGCCGTTCTG 77

Qy   1829 TAAAACTTCATTACTCTTTGGCCCATTTGTATGCACGATGTTTGTTTTTAAAACATGGTA-1887
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 TAAAACTTCATACCTCTTTGGCCATTTGTATGCATGATGTTTGGTTTTTAAAACATGGAAC 17

Qy   1888 TAATGAAATTGTGTACT 1903
              |||||||||||||||||
Db   16 TACTGAAATTGTGTACT 1

RESULT 13
LOCUS   BG698557
DEFINITION BG698557 602650594F2 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4801702 5',
            mRNA sequence.
ACCESSION   BG698557
VERSION   BG698557.1 GI:13965967
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 733)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10694 row: l column: 23
            High quality sequence start: 33
            High quality sequence stop: 731.
FEATURES   Location/Qualifiers
     source   1..733
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4801702"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

BG698557   733 bp   mRNA   linear   EST 07-MAY-2001

ORIGIN

Best Local Similarity   98.3%;   Pred. No. 5.2e-90;
Query Match   32.2%;   Score 644.2;   DB 12;   Length 733;
Matches 683;   Conservative   0;   Mismatches   8;   Indels   4;   Gaps   3;

```
Qy     1198 CCCGGGGTCTCGCTCTTGATAAAAATAACACGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 1257
                ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       40 CCCGGGGTCTCGCTTCTTGATAAAATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGAT 99
Qy     1258 TCAAGAACTGGTGTGGCACCAGATGTCTTCCGCTGAAAACATGAAGTCAGAGAGGACACC 1317
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      100 TCAAGAACTGGTGTGGCACCAGATGTCTTCCGCTGAAAACATGAAGTCAGAGAGGACACC 159
Qy     1318 CATCATTTGGTCAACTATATTCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGG 1377
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      160 CATCATTTGGTCAACTATATTCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGG 219
Qy     1378 TACTTAGTTAGGGTGGATGAAAAAGCACAGAAGAAGAACTTACGCTCCGGCGTCCCACTCCCAAAT 1437
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      220 TACTTAGTTGAGGGTGGATGAAAAAGCACAGAAGAAGAACTTACGCTCCGGCGTCCCACTCCCAAAT 279
Qy     1438 GATGACACGTTTGGATGAAGGAGTAGGACTTGTGCATTCAAAACATAGCAACTGAGCATATT 1497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      280 GATGACACGTTTGGATGAAGGAGTAGGACTTGTGCATTCAAAACATAGCAACTGAGCATATT 339
Qy     1498 CCAAGCCCTGCCAAGAAAAGTACCAAGACTCCCTGCCCACAGCAGCAGAACCTGAAGCAGCT 1557
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      340 CCAAGCCCTGCCAAGAAAAGTACCAAGACTCCCTGCCCACAGCAGCAGAACCTGAAGCAGCT 399
Qy     1558 GTCATTAGTTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTG 1617
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      400 GTCATTAGTTAATGGGGAACATTAAGATACTCTGTGAGGTGCAAGACTTCAGGGTGGGGTG 459
Qy     1618 GGCATGGGGTG-GGGGTATGGGGAACAGTTGGAGGAATGGGGGAATATCTCGGGGGATAATTTAA 1676
                ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      460 GGCATGGGGTGCGGGGTATGGGGAACAGTTGGAGGAATGGGGGAATATCTCGGGGGATAATTTAA 519
Qy     1677 A-GGATTACATGTTGTTATGTGAATTTTTATGTGACTGAGGAGCCTGGATGACTATCGTG 1735
               | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      520 ACGGATTACATGTTGTTATGTGAATTTTTATGTGACTGAGGAGCCTGGATGACTATCGTG 579
Qy     1736 TACTTGGGAAAGTCTCTTTTGCTCTATTTGCTTGACATGCTTCCTGTGTTGGTCTGGCCAAT 1795
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db      580 TACTTGGGAAAGTCTCTTTTGCTCTATTTGCTTGACATGCTTCCTGTGTTGGTCTGGACAAT 639
Qy     1796 GCCAAATGTACTCGAATGAATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATTT 1855
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      640 GCCAAATGTACTCGAATGAATGTTAAGGGCTCTGTAAAACTTCATACCTCTTTGGCCATAT 699
Qy     1856 GTATGCACATGATTGTTGGTTTTTTTAAACATGGTATAA 1890
                |||||||||||||||||||||||||||||||||||||||
Db      700 GTATGCACATGATTGTTGGTTTTTTTAAACATGGTATAA 732
```

ORIGIN

Best Local Similarity   98.3%;   Pred. No. 6.3e-90;
Query Match   32.1%;   Score 643.6;   DB 14;   Length 789;
Matches 697;   Conservative   0;   Mismatches   92;   Indels   0;   Gaps   0;

```
Qy      710 GCTGCTACTCTGTCTACTGCAAAAAAGATCCATGCCCAGTGGCAGAAAGAGGGAAATGATA 769
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GCTGCTATTCTGTCTACTACCGCAAAAAAGATCCGGCCCCAGTGGCAGAAAGAGGGAAATGATA 60
Qy      770 AAGATTTTACCTTGAATGATTTTGGCTTCATGATCTTTCATGAGTCTTCACTACCATATTGTAAACTGG 829
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 AAGATTTTACCCTGAATGATTTTGGCTTCATGATCTTTCACTCGCCATACTGTAAACTGG 120
Qy      830 TTCAGAAAATCTTACCTCGGATCGTTGCTGAATGACTTCCTTAATGACCAGAATAGAGATA 889
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 TGCAGAAAATCTTAGCTACGGATGGTTCCTGAAATGACTTCCTTAATCATCAAAAACAGAGACA 180
Qy      890 AAAATAGTATCTATAGTGGCCTCGGAAGCCTTTGGGGATTGGTTAAATTAGAAGACACCTACT 949
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AAAACAGTATTTACAGTGGCTGGCTCGGAAGCCTTTGGGGATTGTTGAAATTAGAAGATACTTACT 240
Qy      950 TTGATAGAGATGTGGAGAGGGCATTATGAAGGCTAGCTTCGTGAACTCTTCAGTCAGAAA 1009
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TCGACAGAGATGTGGAAAGGCATTTTATGAAGGCTAGTGTGCTGAGCTATTCAACCAGAAA 300
Qy     1010 CAAAGGCATCTTTACTTGTGTATCAAATCAAAATGGAAATAATGTACACATTCAGTATATG 1069
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 CAAAGGCATCTTTGCTTGTGTATCGGAATCAAAATGAAACACATGTACACATCCTCTGTATACG 360
Qy     1070 GTTCCCTTGCATCTTCTTCTAGCACAGTACTCACCTCAGCAATTAGCAGGGAAGGAAATTG 1129
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 15

BX419944

```
LOCUS       BX419944              1201 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX419944 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE18 5-PRIME, mRNA sequence.
ACCESSION   BX419944
VERSION     BX419944.1  GI:30650862
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3098.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022BC09QP1&cluster=3098.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022BC09QP1.
FEATURES             Location/Qualifiers
     source          1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE18"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN
```

```
Query Match           31.3%;  Score 626;  DB 13;  Length 1201;
Best Local Similarity 84.0%;  Pred. No. 2.7e-87;
Matches   818;  Conservative   4;  Mismatches  19;  Indels  133;  Gaps   5;
```

```
Qy    100 ACTCCCTTTCCTCTGCGCGCTCGGTCACGCTTGCTCTTCTTCACTGATCACTT 159
          |||||||||||| |||| || ||  || ||| |||||||||| ||||| |||||
Db     45 ATTCCCTTTCCTCTCCGCTGCGGTCACGCTTGCTCTTTTCACCATGCCGGATCACTT 104

Qy    160 CCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGGGGAATTGTTGCCCTTGAGATCTAT 219
          |||||||||||||||  ||| ||||||||||| ||||||||||||||||||| |||||||
Db    105 CCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGGGGAATTGTTGCCCTTGAGATCTAT 164

Qy    220 TTTCCTTCTCAATATGTTGATCAAGCAGAAGCTTGGAAAAAAATATTGATGCTGGA 279
          ||| |||||||||||||||||| |||||||||||| |||||||||||||||||||||
Db    165 TTTCCTTCTCAATATGTTGATCAAGCAGAAGCTTGGAAAAAATATGGTGTGATGCTGGA 224

Qy    280 AAGTATACCAATTGGCTTGGGCCCCAGCGCCACAGATGGGCCTTCTGCAGATAGGAGATATT 339
          || ||||||||||||||||||||||| |||||| |||||||||  |||||||||||||||
Db    225 AAGTATACCAATTGGCTTGGGCCCCAGCGCCACAGATGGGCCTTCTGCAGATAGGAGAGATATT 284

Qy    340 AACTCTCTTTGCATGACTGGTGGTTCAGAATCTTATGGAGAGAAATAACCTTTCCTATGAT 399
          ||||| |||||||||||||||| ||||| |||||||||||| |||||||||| |||||||
Db    285 AACTCTCTTTGCATGACTGGTTCAGAATCTTATGGAGAGAAATAACCTTTCCTATGAT 344

Qy    400 TGCAATTGGGGGCCTGGAAGTTGGACAGAGAACAATCATCAGACAAATCAAAGTCTCTGAAG 459
          ||||||||||| ||||||||  |||||||||| ||||||||||||||||||||||| ||||
Db    345 TGCATTGGGGGGCCTGGAAGTTGGACAGAGAACAATCATCAGACAAATCAAAGTCTGTGAAG 404

Qy    460 ACTAATTTGATGCACGCTGTTTGAAGAGTCTTGGGAATACAGATATGAAGGAAATCACACA 519
          |||||||||||||| ||||||||||||||| |||||||||||||||||||||||||||||
Db    405 ACTAATTTGATGCACGCTGTTTGAAGAGTCTTGGGAATACAGATATGAAGGAAATCACACA 464

Qy    520 ACTAATGCATGCTATGGAGAGGCACAGCTGCTGTCTGTCTTCAAATGCTTGTTAAACTGGATTGAGTCC 579
          |||||||||||||||||||||||||| |||||| ||||||||| |||||| |||||||||||||||||
Db    465 ACTAATGCATGCTATGGAGAGGCACAGCTGCTGTCTGTCTTCAAATGCTTGTTAAACTGGATTGAGTCC 524

Qy    580 AGCTCTTGGGAT------------------------------------ 591
          ||||||||||||
Db    525 AGCTCTTGGGATGGACGGGTATGCCCCTGGTAGTTGCAGGAGTATTGCTGTATATGGCCACA 584

Qy    592 ---------------------------------GGGCTTCGTGGGACCACACATATGCAACATGCCTATTGATTTTTAC 633
                                           |||||||||||||||||||||||||||||||||||||| ||||
Db    585 GGAAATGCTAGACCTACAGGTGGAGCAGTGCTTGGGCTTCGTGGGACCACACATATGCAACATGCCTATTGATTTTTAC 644

Qy    645 CCTTTAAATTTTTGAACGAGGGCTTCGTGGGCACAGATGTAGCACATGCCTATGATTTTTAC 704
          ||||||||||||||||||||||||||||  ||||||||||| |||||||||| |||||||||
Db    645 CCTTTAAATTTTTGAACGAGGGCTTCGTGGGCACAGATGTAGCACATGCCTATGATTTTTAC 704

Qy    694 CTCAGTTGCATTAGACCGGCTCGCCTACTCCTGTCTTACTCGCAAAAGATCCATGCCCAGTGGCAG 753
          || |||||||||| ||||||||| ||||||||||||||| ||||||||||||||||||||||||||
Db    705 AAGCCTGATATGCTATCTGAATATCCTATAGTAGAATGGAAAACTCCATACACAGTCAGTGCTAC 764

Qy    754 AAAGAGGGAAATGATAAAGA---TTTTACCTTGAAATGATTTGGGCTTCATGATCTTTCACT 812
          |||||| |||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db    765 CTCAGTTGCATTAGACCGGCTCGCCTACTCCTGTCTTACTCGCAAAAGATCCATGCCCAGTGGCAG 824

Qy    825 AAAGAGGGAAATGATAAAGATTTTTACCTTGAAATGATTTGGGCTTCATGATCTTTCACT 884
          ||||||||||||||||||||| | |||||||||||||||||||||||||||||||||||
Db    813 ACCATATTGTAAACTGGTTCAGAAATCTCTAGCCTCGGATGTTGCTGAAATGAC-TCCTTAA 872

Qy    873 TGACCAGAATAGAGAGAGTAAAAATAGTATCTATAGGTGGCCTGGAAGCCTTGGGATGGTTAA 932
          ||||| ||| ||||||||| ||||||||||||||||||||| ||||||||  |||||||||||
Db    885 ACCATATTGTAAACTGGTTCAGAAATCTCTAGCCTCGGATGTTGCTGAAATGAC-TCCTTAA 943

Qy    933 ATTAGAAGACACCT 946
          ||  |||||| ||
Db    944 TGMCAGATAGAGAGAGAAAAAATAGTA---CTATAGTGGCCCTGG-AGCCTTTGGGGAGTAA 998
```

```
Qy    361 GTTCCCTTGCTTCTGTTCTTCTGGCACAGTAGTACTCACCTCACACAGTTGGCCGGGAGAGGAATTG 420
          |||||||||| |||| | |||||| |||||   |||||||||||||||||| ||||||||||||||
Qy   1130 GAGTGTTTTTCTTATGGGTTCTGGGTTTGGCTGCCACACTCTTGTACTTCTTCTTAAAGTCACACAAG 1189
          |||||| |||| || |||||||||||||||| ||||||| |||  ||||| ||||||||||||||||
Db    421 GAGTGTTCTTACGGTTCTGGGTTTGGCTGCCACACTCTACTCCCTTAAAGTCACACAAG 480

Qy   1190 ATGCTACACCGGGGTCTGCTTCTTGATAAAATACAGCAAAGTTATGTGATCTTAAATCAA 1249
          ||||| |||||| | |||||||||| |||||||  ||||||||||||||| |||||| |||
Db    481 ATGCCCACACCAGGATCTGCTTCTTGACAAAATAACAGCAAAGTTTATGTGACCTTAAGTCAA 540

Qy   1250 GGCTTGATTCAAGAACTGGTGTGGCCACCAGAGTGTCTTCCGCTGAAAACATGAAGCTCAGAG 1309
          |||||||| |||||| ||||| |||||||| ||||||||||||||||||||||||||||||||
Db    541 GGCTTGACTCAAGAACGGTGTCGTGGCCACCCAGAGTGTCTTCCGCTGAAAACATGAAGCTCAGAG 600

Qy   1310 AGGACACCCATCATTTGGTCAACTATATTCCCAGGGGTTCAAATAGATTCACTCTTTGAAG 1369
          |||||||| |||||||||||||| || |||| ||||||||  |||||||||||||||||||
Db    601 AGGACACACATCACTTAGCCAACTATAATTCGCCAGGTTCAAATAGATTCACTCTTTGAAG 660

Qy   1370 GAACGTGGTGTACTTAGTTAGGGTGGATGGAAAAGCACAGAGAAGAACTTACCGTCGGGGTCCCA 1429
          ||||  | || ||||  || |||||||| ||||||||||||| |||||||||| ||| ||||||
Db    661 GAACATGGTATCTCAGTGGACTGATGAAAAGCACAGAGAAGAACTTACCGCCGCGGGTCCCT 720

Qy   1430 CTCCAAATGAATGACACTTTGGAATGGAAGGAGGAGGAGTAGGACTTGTCATTCANACATAGCAACTG 1489
          || |||||||||| |||||| || ||||||  |||||| ||||| |||||||||| || || |||||
Db    721 CCACANNATGACACAGTTGGGATGAAAGGAGGTGGGGACTTGTGCATTCANACACAGCTACAG 780

Qy   1490 AGCATATTC 1498
          |||||||||
Db    781 AGCATATTC 789
```

Db   999 TWAGARGCACCTCT 1012

Search completed: June 24, 2004, 11:14:15
Job time : 4695 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 13:22:38 ; Search time 4601 Seconds
                 (without alignments)
                 4502.928 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV..........PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODEL=frame+_p2n.model
-Q=/cgn2_1/USPTO_spool_p/USI0622516/runat_23062004_162633_648/app_query.fasta_1.647
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=USI0622516 @CGN_1_1_5600 @runat_23062004_162633_648 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY_NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl

1:	gb_ba:*	29:	em_vi:*
2:	gb_htg:*	30:	em_hum:*
3:	gb_in:*	31:	em_htg_inv:*
4:	gb_om:*	32:	em_htg_other:*
5:	gb_ov:*	33:	em_htg_mus:*
6:	gb_ov:*	34:	em_htg_pln:*
7:	gb_pat:*	35:	em_htg_rod:*
8:	gb_ph:*	36:	em_htg_mam:*
9:	gb_pl:*	37:	em_htg_vrt:*
10:	gb_pri:*	38:	em_sy:*
11:	gb_ro:*	39:	em_htgo_hum:*
12:	gb_sts:*	40:	em_htgo_mus:*
13:	gb_sy:*	41:	em_htgo_other:*
14:	gb_un:*		
15:	gb_vi:*		
16:	gb_ba:*		
17:	em_hum:*		
18:	em_in:*		
19:	em_mu:*		
20:	em_or:*		
21:	em_om:*		
22:	em_ov:*		
23:	em_ph:*		
24:	em_pl:*		
25:	em_pr:*		
26:	em_ro:*		
27:	em_sts:*		
28:	em_un:*		

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2511	100.0	2002	6	AR399482	AR399482 Sequence
2	2480	98.8	1563	10	BT007302	BT007302 Homo sapi
3	2480	98.8	1650	9	HUM3H3M	L25798 Homo sapien
4	2480	98.8	2068	9	BC000297	BC000297 Homo sapi
5	2439	97.1	1685	9	HSCOAS	X66435 H.sapiens m
6	2403.5	95.7	3195	9	AK095492	AK095492 Homo sapi
7	2362	94.1	3275	6	AX700129	AX700129 Sequence
8	2362	94.1	3275	6	AX827303	AX827303 Sequence
9	2362	94.1	3275	10	RNCHMCA	X55625 Rattus norv
10	2348	93.5	2703	10	BC023851	BC023851 Mus muscu
11	2348	93.5	3202	10	BC034317	BC034317 Mus muscu
12	2348	93.5	3260	10	BC029693	BC029693 Mus muscu
13	2344	93.3	3278	10	BC031363	BC031363 Mus muscu
14	2097	83.5	1874	5	CHKHMGCOAS	M60657 Chicken HMG
15	2089	83.2	1824	9	I65510	I65510 Sequence 1
16	2003	79.8	3098	5	BC042829	BC042829 Xenopus 1
17	1957.5	78.0	3205	5	BC049456	BC049456 Danio rer
18	1691	67.3	127832	9	AL356361	AL356361 Human DNA
19	1691	67.3	170834	2	AC021792	AC021792 Homo sapi
20	1513	60.3	2058	6	AX337379	AX337379 Sequence
21	1513	60.3	2058	6	AX411166	AX411166 Sequence
22	1513	60.3	2058	9	HSHMGCOAS	X83618 H.sapiens m
23	1513	60.3	2098	6	BC044217	BC044217 Homo sapi
24	1513	60.2	1994	6	AX163779	AX163779 Sequence
25	1511.5	60.2	1994	10	RATHMGCOA	M33648 Rat mitocho
26	1494	59.5	1991	10	AK098104	AK098104 Mus muscu
27	1493	59.5	1994	10	AC017201	AC017201 Drosophil
28	1493	59.5	2011	3	BC014714	BC014714 Mus muscu
29	1466	58.4	1744	4	U90884	U90884 Sus scrofa
30	1396	55.6	163423	2	AC123738	AC123738 Mus muscu
31	1396	55.4	144405	9	AC114947	AC114947 Homo sapi
32	1362.5	54.3	142657	2	AC008773	AC008773 Homo sapi
33	1359	54.1	1658	9	BGHMG	BGHMG Sequence
34	1353	53.9	1296	10	MMU12790	U12790 Mus musculu
35	1337	53.2	2588	3	AY051743	AY051743 Drosophil
36	1337	53.2	97346	2	AC017201	AC017201 Drosophil
37	1337	53.2	157835	3	AC009356	AC009356 Drosophil
38	1337	53.2	159640	3	AC002230	AC002230 Drosophil
39	1337	53.2	258223	3	AE003807	AE003807 Drosophil
40	1332	53.0	236502	2	AC098186	AC098186 Rattus no
41	1272.5	50.7	1716	3	BGHMGCOA3	BGHMGCOA3 Sequence
42	1223.5	48.7	1881	3	AK112832	AK112832 B.germanica
43	1204.5	48.0	259650	2	AC107176	AC107176 Rattus no
44	1179	47.0	2015	3	AF160002	AF160002 Dendrocto
45	1115	44.4	1542	9	U12789	U12789 Human clone

RESULT 1

ALIGNMENTS

RESULT 1

```
AR399482
LOCUS       AR399482     2002 bp    DNA     linear   PAT 18-DEC-2003
DEFINITION  Sequence 1 from patent US 6620608.
ACCESSION   AR399482
VERSION     AR399482.1  GI:40141523
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2002)
AUTHORS     Gong,F., Yan,C., Di Francesco,V. and Beasley,E.M.
TITLE       Isolated human synthase proteins
JOURNAL     Patent: US 6620608-A 1 16-SEP-2003;
FEATURES             Location/Qualifiers
     source          1..2002
                     /organism="unknown"
                     /mol_type="genomic DNA"
ORIGIN
```

US-10-622-516-2 (1-478) x AR399482 (1-2002)

```
Alignment Scores:
Pred. No.:
Score:                 4.42e-205
Score:                 2511.00          Length:        2002
Percent Similarity:    100.00%          Matches:        478
Best Similarity:       100.00%          Conservative:     0
Best Local Similarity: 100.00%          Mismatches:       0
Query Match:           6                Indels:           0
DB:                    6                Gaps:             0
```

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCCAAAAGATGTGGGAATTGTT 204

Qy     21 AlaLeuGluIleTyrPheProSerGlnTyrValAlaGluLeuGluLysTyrAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGCAGAGTTGGAAAAATATGAT 264

Qy     41 GlyValAspProAlaAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    265 GGTGTAGATCCTGCAGCAGGAAAGTATACAATTGGCCTTGGCCAAGCCAAAATGGGCTTCTGCACA 324

Qy     61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 GATAGAGAAGATATTAACTCTTTGTGCATGACTGTTGTTCAGAATCTTATGGAGAGAAAT 384

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 AACCTTTCCTATGATTGCATTGGCCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAA 444

Qy    101 SerLysSerValLysSerThrArgLeuMetGlnLeuSerGlyValAsnThrAspIle 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    445 TCAAAGTCTGTGAAGAGCACTAGATTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 504

Qy    121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    505 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTCTGTTCAATGCTGTT 564

Qy    141 AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyTyrThrHisMetGlnAlaTyr 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    565 AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGCTACACATATGCAACATGCCTAT 624

Qy    161 AspProSerTyrGlyProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 GATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTGGATGGAAAACTCTCCATA 684

Qy    181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    685 CAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATCCATGCC 744

Qy    201 GlnIleTrpGlnIleLysAspGlyAsnAspPheTheThrLeuAsnAspPheGlyPheMetIle 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 CAGTGGCAGAAGAGGGAAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTCATGATC 804

Qy    221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLysAsnAsp 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 TTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGAATGCTGAAATGAC 864

Qy    241 PheLeuAsnAspGlnAsnArgAspPheLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    865 TTCCTTAATGACCAGAATAGAGATTTAAAAAATAGTATCTATAGTGGGCTGGAAGCCTTTGGG 924

Qy    261 AspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    925 GATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATGAAGGCT 984

Qy    281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985 AGCTCTGAACTCTTCAGTCAAAAGACAAAAGCATCTTTACTTGTATCAAATCAAAATGGA 1044

Qy    301 AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045 AATATGTACACATCTTCAGTATATGGGTCTCTTGCATCTGTTCTAGCACAGTACTCACCT 1104

Qy    321 GlnGlnLeuGlnAlaGlyLysLysAspLeuGluValPheSerTyrGlySerGlyLeuAlaThr 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105 CAGCAATTAGCAGGAAAGAAAGATTTGGAAGTGTTTTCTTATGGGTCTGGTCTTGCCACT 1164

Qy    341 LeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspAspLysLysIleThr 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165 CTGTACTCTCTTGTAACGCAAGATGCTACACCCGGGTCTGCTCTTGATGATAAAAAAATAACA 1224

Qy    361 AlaSerLeuCysAspAlaLysTyrAlaAspSerArgLeuAspAspArgThrGlyValValValGly 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225 GCAAGTTTATGTGATGCTAAATATGCAGACAGTCGTTTGGACGATAGAACTGGTGTGGTTGTTGGA 1284

Qy    381 PheGluAsnMetLysLeuAspGluHisTyrAlaThrArgIleGluProGln 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285 TTCGGTGAAAACATGAAACTAGATGAACACTATGCCACTCGAATAGAACCCCAG 1344

Qy    401 GlySerIleAspGluSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluLysHis 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345 GGTTCAATAGATGAATCACTCTTTGAAGGAACGTGGTACCTTGTGAGAGTAGATGAAAAGCAC 1404

Qy    421 ArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlyIleGlyGly 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1405 AGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGGAATAGGAGGA 1464

Qy    441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysValProArg 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465 CTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTACCAAGA 1524

Qy    461 LeuProAlaThrAlaAlaLeuProGlyIleAlaValAlaIleSerAsnGlyGluHis 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1525 CTCCCTGCCACAGCAGCACTGCCAGGAATTGCAGTTGCCATTAGTAATGGGGAACAT 1578
```

RESULT 2

```
BT007302
LOCUS       BT007302     1563 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
            (soluble) mRNA, complete cds.
ACCESSION   BT007302
VERSION     BT007302.1  GI:30583442
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1563)
AUTHORS     Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
TITLE       Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 1563)
AUTHORS     Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
```

TITLE　　　　Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
　　　　　　　Phelan,M. and Farmer,A.
JOURNAL　　　Direct Submission
　　　　　　　Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
　　　　　　　Circle, Palo Alto, CA 94303, USA
COMMENT　　　This CDS clone is a part of a collection of human full length
　　　　　　　expression clones generated by BD Biosciences Clontech and the
　　　　　　　Harvard Institute of Proteomics. Each CDS has been cloned in two
　　　　　　　forms: with and without stop-codon (to allow fusion with C-terminal
　　　　　　　tag). The CDS has been directionally cloned using BD In-Fusion(TM)
　　　　　　　cloning system between the SalI and HindIII sites of the pDNR-DUAL
　　　　　　　vector. Additional sequences in the clone: 'ACC' after SalI site
　　　　　　　and before 'ATG' to provide Kozak consensus sequence; 'GG' after
　　　　　　　last codon and before HindIII site to maintain reading frame.
　　　　　　　Clone distribution: http://bioinfo.clontech.com/orclones.

FEATURES　　　　　　　　Location/Qualifiers
　　source　　　　　　1..1563
　　　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　　　/mol_type="mRNA"
　　　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　　　/clone_lib="BD Creator(TM) CDS Library derived from MGC
　　　　　　　　　　　collection"
　　　　　　　　　　　/clone="GH0878X1.0"
　　　　　　　　　　　/lab_host="DH5alpha T1 resistant"
　　　　　　　　　　　/note="Vector: pDNR-Dual"
　　CDS　　　　　　　1..1563
　　　　　　　　　　　/codon_start=1
　　　　　　　　　　　/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
　　　　　　　　　　　(soluble)"
　　　　　　　　　　　/protein_id="AAP35966.1"
　　　　　　　　　　　/db_xref="GI:30583443."
　　　　　　　　　　　/translation="MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDA
　　　　　　　　　　　GKYTIGLGQAKMGFCTDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
　　　　　　　　　　　SVKTNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
　　　　　　　　　　　AVYATGNARPTGGVGAVALLIGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDG
　　　　　　　　　　　KLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLA
　　　　　　　　　　　RMLNDFLNDNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELPSQKTKAS
　　　　　　　　　　　LLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
　　　　　　　　　　　TPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFE
　　　　　　　　　　　GTWYLVRVDEKHRRTYARRPPTPNDDTLDEGVGLVHSNIATEHIPSPAKKVPRLPATPA
　　　　　　　　　　　EPEAAVISNGEH"

ORIGIN

US-10-622-516-2 (1-478) x BT007302 (1-1563)

Alignment Scores:
Pred. No.:　　　　　1.42e-202
Score:　　　　　　　2480.00
Percent Similarity:　91.92%
Percent Identity:　　91.92%
Best Local Similarity:　98.77%
Query Match:　　　　9
DB:

Db：US-10-622-516-2 (1-478)　x　BT007302 (1-1563)

Length:	1563					
Matches:	478					
Mismatches:	0					
Conservative:	0					
Indels:	42					
Gaps:	1					

Qy　　1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　101 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAGAGATGTGGGAATTGTT 160

Qy　　1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　101 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAGAGATGTGGGAATTGTT 160

Qy　101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db　301 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 360

Qy　21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db　61 GCCCTTGAGATCTATTTTCCTTCAATATGTTGATCAAGCAGAGTTGGAAAATATGAT 120

Qy　121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db　361 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT 420

Qy　41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
Db　121 GGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGACAAGCTAAGATGGGCTTCTGCACA 180

Qy　141 AsnThrIleGluSerSerSerTrpAsp------------------------------- 149
Db　421 AACTGGATTGAGTCCAGCTCTTGGGGATATGGACCGGTATGCCCTGGTAGTTGCAGGAGAT 480

Qy　61 AspArgArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db　181 GATAGAGAAGAGATAAATAACTCTTTGTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 240

Qy　149 -------------GlyLeuArgGlyThrHisMetGlnHis 158
Db　481 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 540

Qy　159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db　541 ATTGGGCCAAATGCTCCTTAATTTTTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 600

Qy　81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
Db　241 AACCTTTCCTATGATTGCATTGGGCGCGGTCTGGAAGTTGGAACAGAGACAATCATCGACAAA 300

Qy　179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db　601 GCCTTATGATTTTTTCACAAGCCTGAATAGTGCTATTCGAATATCCTATAGAGAGTAGATT 660

Qy　199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db　661 TCCATACAGTGCTACCTCAGTGCATTAGACCGGCTCTACTCGTCTTAGTCGCAAAAAGATC 720

Qy　219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuAsn 238
Db　721 CATGCCCAGTGGCAGAAAGAGGGAAATGATAAGGATTTTACCTTGAATGATTTTGGCTTC 780

Qy　239 AspPheLeuAsnAspAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
Db　781 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 840

Qy　259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db　841 AATGACTTCCTTAATGACAACAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC 900

Qy　279 LysAlaSerSerGluLeuProSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db　901 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 960

Qy　299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db　961 AAGGCTAGCTCTGAACTCCCTTCAGTCAGAAAACAAAGGCAATCTTTACTTGTTATCAAATCAA 1020

Qy　319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db　1021 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1080

Qy　339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db　1081 TCACCTCAGCAAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTATAGTGGTTCGGTTGGCT 1140

Qy　359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db　1141 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACCCCGGGTTCTGCTCTTGATAAA 1200

Qy　379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db　1201 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA 1260

Qy　399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db　1261 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGACACCCATCATTTGGTTCAACTATATT 1320

Qy　419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyVal 438
Db　1321 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTTACTTAGGTTAGGGTGATGAA 1380

```
RESULT 3
HUM3H3M
LOCUS       HUM3H3M                 1650 bp    mRNA    linear   PRI 24-FEB-1995
DEFINITION  Homo sapiens 3-hydroxy-3-methylglutaryl coenzyme A synthase,
            complete cds.
ACCESSION   L25798
VERSION     L25798.1  GI:410027
KEYWORDS    3-hydroxy-3-methylglutaryl coenzyme A synthase.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1650)
AUTHORS     Rokosz,L.L., Boulton,D.A., Butkiewicz,E.A., Sanyal,G., Cueto,M.A.,
            Lachance,P.A. and Hermes,J.D.
TITLE       Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase:
            expression, purification, and characterization of recombinant
            wild-type and Cys129 mutant enzymes
JOURNAL     Arch. Biochem. Biophys. 312 (1), 1-13 (1994)
MEDLINE     94304197
PUBMED      7913009
COMMENT     Original source text: Homo sapiens fetal adrenal cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1..1650
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="fetal adrenal"
     CDS             22..1584
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl coenzyme A synthase"
                     /protein_id="AAA62411.1"
                     /db_xref="GI:410028"
                     /translation="MPGSLPLNAEACWPKDGIVALEIYPSQYVDQAELEKYDGVDA
                     GKYTIGLGQAKMGFCDTDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
                     SVKTNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
                     AVYATGNARPTGGVGAVALLIGPNAPLIFERGLRGTHMQAHYDFYKPDMLSEYIPVDG
                     KLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKAS
                     LLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTGVAPVDFAENMKLREDTHHLVNYIPQGSIDSLFE
                     GTWYLVRVDEKHRRTYARRPTPNDDTLDEGVLVHSNIATEHIPSPAKKVPRLPATRA
                     EPEAAVISNGEH"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.52e-202
Score:                    2480.00
Percent Similarity:       91.92%
Best Local Similarity:    91.92%
Query Match:              9
DB:
```

US-10-622-516-2 (1-478) x HUM3H3M (1-1650)

```
Length:          1650
Matches:          478
Conservative:       0
Mismatches:        42
Indels:             0
Gaps:               1
```

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspGlyIleValAla   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCCCGGATCACTTCCTTGAATGCAAGAGCTTGGCCTGCCAAAAGATGGGGATTGT   81

Qy     21 AlaLeuGluIleTyrProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 GCCCTTGAGATCTATTTTCCTTCTCAAATAGTTGATCAAGCAGAGTTGGAAAAATATGAT  141

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 GGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCCAAGCCAAGATGGGGTTCTGCACA  201

Qy     61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT  261

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAA  321

Qy    101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA  381

Qy    121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCCGCTGTTTTCAATGCTGTT  441

Qy    141 AsnTrpIleGluSerSerSerTrpAsp--------------------------------  149
          ||||||||||||||||||||||||||||
Db    442 AACTGGATAGAGTCCAGCTCTTGGGACGGGTATGCCCTGGTAGTTGCAGGAGATATT  501

Qy    149 ----------------GlyLeuArgGlyThrHisMetGlnHis  158
                          ||||||||||||||||||||||||||
Db    502 GCTGTATATGCCACAGGAAATCTAGACCTTACAGGTGGAGTGGGAGCAGTAGCTCTGCTA  561

Qy    150 ------------------------------------------  150
Db    562 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGGGACACATATGCAGCATTACGAAACAT  621

Qy    159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu  178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATTGTAGATGGGAAACTC  681

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle  198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGTTACTCTGTCTACTGCAAAAAGATC  741

Qy    199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe  218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 CATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCCTGAATGATTTTGGGCTTC  801

Qy    219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu  238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAGTCTCTAGCTCGGATGTTGCTG  861

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC  921

Qy    259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet  278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG  981

Qy    279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln  298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    982 AAGGCTAGCTCCGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTTAGCAATCAA  1041

Qy    299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr  318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTGGCTTCTGTTCTTGCACAGTAC  1101

Qy    319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla  338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1102 TCACCTCAGCAAATTGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGGTCTGGTCTTGCT  1161

Qy    339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1162 GCCACTCTGTACTCTCTTAAAGTCACAGATGCTACACCGGGGTCTGCTCTTGATAAA  1221
```

RESULT 4

BC000297

LOCUS   BC000297   2068 bp   mRNA   linear   PRI 30-SEP-2003
DEFINITION   Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1
             (soluble), mRNA (cDNA clone IMAGE:2819708), partial cds.
ACCESSION   BC000297
VERSION   BC000297.2   GI:33991030
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2068)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2068)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   On Aug 20, 2003 this sequence version replaced gi:12653064.
             Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory

             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC),
             Gaithersburg, Maryland;
             Web site: http://www.nisc.nih.gov/
             Contact: nisc_mgc@nhgri.nih.gov
             Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
             Blakesley,R.W. Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
             Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
             Hansen,N.L., Ho,S-L., Karlins,E., Kwong,P.N., Laric,P., Legaspi,R.,
             Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
             McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
             Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
             Young,A., Zhang,L.-H. and Green,E.D.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAL Plate: 1 Row: a Column: 23
             This clone was selected for full length sequencing because it
             passed the following selection criteria: matched mRNA gi: 4504428.

FEATURES   Location/Qualifiers
     source   1..2068
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2819708"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene   <1..2068
                     /gene="HMGCS1"
                     /note="synonym: HMGCS"
                     /db_xref="LocusID:3157"
                     /db_xref="MIM:142940"
     CDS   <1..1668
                     /gene="HMGCS1"
                     /codon_start=1
                     /product="HMGCS1 protein"
                     /protein_id="AAH00297.2"
                     /db_xref="GI:33991031"
                     /db_xref="LocusID:3157"
                     /db_xref="MIM:142940"
                     /translation="HSLSSAARSRLCPKEETVTDLETAVLYPSHSSFTMPGSLPLNA
                     EACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINS
                     LCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT
                     TNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTEGVGAVALLIGP
                     NAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIH
                     AQWQKEGNDKDFLNDFGPMIFHSPYCKLVQKSLARMLLNDFLNDQNDRDKNSIYSGLE
                     AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVL
                     AQYSPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSR
                     TGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPN
                     DDTLDEGVLVHSNIATEHIPSPAKKVRPLPATAEPERAAVISNGEEH"
     misc_feature   142..1512
                     /gene="HMGCS1"
                     /note="HMG_CoA_synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CDD:pfam01154"

ORIGIN

Alignment Scores:
Pred. No.:   2.07e-202
Score: 2480.00
Percent Similarity: 91.92%
Best Local Similarity: 91.92%
Query Match: 9

                                              Length: 2068
                                              Matches: 478
                                              Conservative: 0
                                              Mismatches: 0
                                              Indels: 42
                                              Gaps: 1

US-10-622-516-2 (1-478) x BC000297 (1-2068)

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 ATGCCTGGATCACTTCCTTTGAATGCAGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 165

Qy   359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA 1281

Qy   379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 GATGTCTTCGCTGAAAACATGAAGCTACGAGAGGACACCCATCATTTGGTCAACTATATT 1341

Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1342 CCCCAGGGGTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1401

Qy   419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1402 AAGCACAGAGAACTTACGCTCGGCGTCCCAATCATGATGACACTTTGGATGAAGGA 1461

Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1462 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAAAAGTA 1521

Qy   459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1522 CCAAGACTCCCTGCCACAGCAGCAGAGCCAGAGGCTGTCATTAGTAATGGGGAACAT 1581

```
Qy    21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp    40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   166 GCCCTTGAGATCTATTTCCTTCCCAGATATGTTGACCAAGCAGAGTTGGAAAAATATGAT   225

Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyInAlaLysMetGlyPheCysThr    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   226 GGTGTAGATGCTGGAAAAGTATACACAATTGGGCTTGGGCCAAGATGGGCCTTCTGCACA   285

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn    80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   286 GATAGAGAAGATATTAACTCTCTTTGCATGACAGTGGTTCAGAATCTTATGGAGAGAAAT   345

Qy    81 AsnLeuSerTyrAspCysIleGlyTyrValGlyGlyThrGluThrIleIleAspLys    100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   346 AACCTTTCCTATGATTGCATTGGGTATGTTGGAGGAACAGAACAATCATCGACAAA    405

Qy   101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   406 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA   465

Qy   121 GlyGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal   140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   466 GGAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT   525

Qy   141 AsnTrpIleGluSerSerSerTrpAsp----------------------------   149
         ||||||||||||||||||||||||||||-------------------------------
Db   526 AACTGGATTGAGTCAAGCTCTTGGGATGGCACGGTAGGTATTGCCCTGGTAGTTGCAGGAGATATT   585

Qy   149 ----------------------------------------------------------   149
         ------------------------------------------------------------
Db   586 GCTGTATATGCCACAGGAAATGTTCTAGACCTACAGGTGGGAGGTCGACAGTAGCTCTGCTA   645

Qy   150 ----------------GlyLeuArgGlyTyrThrHisMetGlnHis   158
         ----------------|||||||||||||||||||||||||||||||
Db   646 ATTGGGCCAAATGCTCTCCTTTAATTTTTGGAACGAGAGGGCCTTCCGTCGCTGGGACACACATATCCAACAT   705

Qy   159 AlaTyrAspPheTyrLysProAspMetLeuSerGlyLeuTyrProIleValAspGlyLeu   178
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   706 GCCTATGATTTTTACAAGCCTGATATGCTATCTGGATTATCCTATAGTAGACGGGCTCCTT   765

Qy   179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrGlySerValTyrCysLysIle   198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   766 TCCATACAGTGCTACCTCAGTGCAATTAGACCGCTGCTACTGTAGCGTCTACTGCAAAAGATC   825

Qy   199 HisIleGlnInTrpIleGlyGlyAsnAspAspSerPheGlyPhe   218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   826 CATGCCCAGTGGATCGGAGGGAATGATAAAGGATTTTTTGGGCTTC   885

Qy   219 MetIlePheHisSerProTyrCysLysGlyLeuValGlnLysGlySerLeuValSerGln   238
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   886 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTTCTAGCTCGGAAGGCATTATG   945

Qy   239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla   258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   946 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC   1005

Qy   259 PheGlyAspValLysLeuGlnAspThrTyrPheAspArgAspValGlyLysAlaPheMet   278
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1006 TTTGGGGATGTTAAATTAGAGACACCTTACTTTGATAGAGATGTGGGAGAAGGCATTATG   1065

Qy   279 LysAlaSerProGlyLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln   298
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1066 AAGGCTAGCCCTGGACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAA   1125

Qy   299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerPheLeuAlaSerValLeuAlaGlnTyr   318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1126 AATGGAAATATGTACACATCTTCAGTATATGGGTCCTTCTTGCATCTGTTCTCAGCAGTAC   1185

Qy   319 SerProGlnIleLeuAlaGlyLysLysArgIleGlyValPheSerTyrGlySerGlyLeuAla   338
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1186 TCACCTCAGCAATTAGCAGGGAAGAAGAATTGGAGTGGTTTTCTTARGGGTTCTGGTTTGGCT   1245
```

```
Qy   339 AlaThrLeuTyrSerLeuLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys   358
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1246 GCCACTCTGTACTCTCTTAAAGTCACAAGATGCTACACCGGGTCTGCTCTTGATAAA   1305

Qy   359 IleThrAlaSerLeuCysAspSerLeuLysSerArgLeuAspSerArgThrGlyValAlaPro   378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1306 ATAACAGCAAGTTATGTGATCAGTCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA   1365

Qy   379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisLeuSerLysValAsnTyrIle   398
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1366 GATGTCTTCGGTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAAGGTTAATTATATT   1425

Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu   418
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1426 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTGAGGGTTGATGAA   1485

Qy   419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly   438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1486 AAGCACAGGAGAACTTACGCCTCGGCGCCACTCCAAATGATGACACTTTGGATGAAGGA   1545

Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleIleGluSerProAlaLysLysVal   458
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1546 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA   1605

Qy   459 ProAlaArgLeuProAlaThrAlaAlaLeuGluProGluAlaAlavalIleLeuSerAsnGlyGluHis   478
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1606 CCAAGACTCCCTGCCACAGCAGCAACCTGAAGCAGCTGTTCATTAGTAATGGGGAACAT   1665
```

RESULT 5
LOCUS   HSCOAS
HSCOAS
LOCUS   HSCOAS   1685 bp   mRNA   linear   PRI 12-SEP-1993
DEFINITION H.sapiens mRNA for HMG-CoA-synthase.
ACCESSION X66435 S48133
VERSION   X66435.1 GI:30008
KEYWORDS   Hydroxymethylglutaryl CoA Synthase.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1685)
  AUTHORS Russ,A.P., Ruzicka,V., Maerz,W., Appelhans,H. and Gross,W.
  TITLE   Amplification and direct sequencing of a cDNA encoding human
           cytosolic 3-hydroxy-3-methylglutaryl-coenzyme A synthase
  JOURNAL Biochim. Biophys. Acta 1132 (3), 329-331 (1992)
  MEDLINE 93041939
  PUBMED   1358203
REFERENCE 2 (bases 1 to 1679)
  AUTHORS Russ,A.
  TITLE   Direct Submission
  JOURNAL Submitted (26-MAY-1992) A. Russ, Labor fur angewandte Biochemie,
           Theodor Stern-Kai 7, W-6000 Frankfurt am Main 70, FRG
  REMARK   revised by [2]
REFERENCE 3 (bases 1 to 1685)
  AUTHORS Russ,A.
  TITLE   Direct Submission
  JOURNAL Submitted (10-AUG-1992) Andreas Russ, Zentrum der biologischen
           Chemie, J.W.-Goethe-Universitaet Frankfurt, Theodor-Stern-Kai 7,
           Frankfurt, 6000, Germany
FEATURES   Location/Qualifiers
     source   1..1685
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HMGCoASynthase"
                     /cell_type="Fibroblast"
     CDS   123..1685
                     /codon_start=1
                     /product="hydroxymethylglutaryl CoA Synthase"
                     /protein_id="CAA47061.1"
                     /db_xref="GI:30009"
                     /db_xref="GOA:Q01581"
                     /db_xref="SWISS-PROT:Q01581"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYPSQYVDQAELEKYDGVDA

ORIGIN

```
GKYTIGLGQAKMGFCTDREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSK
SVKTNLMQLFEESGNTDIEGIDTTNACYGGTAAVFRNVAVNWIESSSWDGRYALVVAGDI
AVYATGNARPTGCVGAVALLIGPNAPLIFERGLRGTHMQHAYDPYKPDMLSEYPIVDG
KLSIQCYLSALDRCYSYYCKKIHAQWQKEANDDFLLNDFGFMIFHSPYCKLVQKSLA
RMLLNDFLNDQNRDKNSIYSGLKAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKAS
LLVSNQNGHNMYTSSVYSSLASVLAQYSPQHLAGKRIGVFSYGSGLAATLYSLKVTQDA
TPGSALDKITASLCDLKSRLDSRTGVAQDVFAENMKLREDTHHLVNYIPQGSIDSLFE
GTWYLVRVDEKHRRTYARRPTPNDDTLDEGVLHSNIATEHIPSPAKKVPRLPATAA
EPEAAVISNGVW"
```

Alignment Scores:

Pred. No.:	5.04e-199
Score:	2439.00
Percent Similarity:	91.12%
Best Local Similarity:	90.93%
Query Match:	97.13%

US-10-622-516-2 (1-478) x HSCOAS (1-1685)

DB:

Length:	1685
Matches:	471
Conservative:	1
Mismatches:	4
Indels:	42
Gaps:	1

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGlnAlaCysTrpProLysAspValGlyIleVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 ATGGCTGGAATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 182

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAlaGluLeuGluLysTyrAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGCAGAGTTGGAAAAATATGAT 242

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 GGTGTAGATGCTGGAAAGTATACCATTGGGCTTGGGCAGGCCAAGATGGGCTTCTGCACA 302

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303 GATAGAGAAGATATTAACTCCTCTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 362

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 AACCTTTCCTATGATTGCATTGGGCGCCTGGAAGTTGGAACAGAGACAATCATCGACAAA 422

Qy  101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  423 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 482

Qy  121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheArgAsnAlaVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483 GAAGGAATGACACAACTAATGCACTGTTCTTCAACTGTGTCTTCAATGCTGTT 542

Qy  141 AsnTrpIleGluSerSerSerTrpAsp----------------- 149
          ||||||||||||||||
Db  543 AACTGGATTGAGTCCAGCTCTTGGGATGCGGTATGCCCTGGTAGTTGCAGGAGATATT 602

Qy  149 ------------------- 149
Db  603 GCTGTATATGCCACAGGAAATGCTAGACGCTACAGGTGGAGTTGGAGCTAGTCTCTA 662

Qy  150 ----GlyLeuArgGlyThrHisMetGlnHis 158
              |||||||||||||||||||||
Db  663 ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCCTTCGTGTGGACACATATGCAACAT 722

Qy  159 AlaTyrAspProTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  723 GCCTATGATCCTTTTACAAGCCTGATATGCTCATTCTGAAATATCCTATAGTAGATGGAAAACTC 782

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  783 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC 842

Qy  199 HisAlaGlnTrpGlnLysGluGlnAspAsnAspPheThrLeuAsnAspPheGlyPhe 218
          |||||||||||||||||||||||||||||||||||||||||||||
Db  843 CATCCCAGTGGCACTGGCAGAAAGAGGCAAATGATAACGATTTTACCTTGAATGATTTTGGCTTC 902
```

```
Qy  219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  903 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 962

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuLysAla 258
          |||||||||||||||||||||||||||||||  :||||||||||||
Db  963 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATATATTCAGGCCTGAAGGCC 1022

Qy  259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1023 TTTGGGGATGTTAAGTTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 1082

Qy  279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1083 AAGGCTAGCTCTGAACTCTTCAGTCAGAAGACTAAAGCATCTTTACTTGTATCAAATCAA 1142

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1143 AATGGAAATATGTACACATCTTCAGTGTATGGTTCCCTTGCATCTGTTCTTGCACAGTAC 1202

Qy  319 SerProGlnHisLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1203 TCACCTCAGCATTTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGGTCTGGTTTGGCT 1262

Qy  339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1263 GCCACTCTGTACTCTCTTAAAGTCACACAGGATGCTACACCGGGGTCTGCTCTTGATAAA 1322

Qy  359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1323 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTTGCACCA 1382

Qy  379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1383 GATGTCTTCGCTGAAAACATGAAGCTTAGAGAGGACACCCATCATTTGGTCAACTATATT 1442

Qy  399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1443 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1502

Qy  419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGlyGly 438
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1503 AAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGA 1562

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1563 GTAGGACTTGTGCATTCAAACATAGCAACAGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1622

Qy  459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGly 476
          |||||||||||||||||||||||||||||||||||||||||||||||
Db 1623 CCAAGACTCCCTGCTACAGCAGCAGAACCTGAAGCAGCAGTTATTAGTAATGGG 1676
```

RESULT 6

AK095492

```
TITLE      NEDO human cDNA sequencing project
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 3195)
AUTHORS    Isogai,T. and Yamamoto,J.
TITLE      Direct Submission
JOURNAL    Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
           Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
           (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT    NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing:
           Research Association for Biotechnology (RAB); cDNA library
           construction: Helix Research Institute (HRI) (supported by Japan
           Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
           HRI, and Biotechnology Center, National Institute of Technology and
           Evaluation; clone selection for full insert sequencing: HRI and
           RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..3195
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PCBBF1000053"
                     /tissue_type="brain"
                     /clone_lib="PCBBF1"
                     /dev_stage="fetal"
                     /note="cloning vector: pME18SFL3"
     CDS             68..1597
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAC04559.1"
                     /db_xref="GI:21754758"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                     GKYTIGLGQAKMGFCTDREDINSLCMTVVQNLMRNNLSYDCIGRLEVGTETIIDKSK
                     SVKNLMQLFEESGNTDIEGIDTTNAVNMIESSSWDGRYAVLVVAGDIAYATGNARPT
                     GGVAAVALLIGNAPLIFERGLRGTHMQHAYDFKPDMLSEYPIVDGKLSIQCYLSAL
                     DRCYSVYCKKIHAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQ
                     NRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMY
                     TSSVYGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITA
                     SLCDLKSRLDSRTGVAPDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEK
                     HRRTYARPTPNDDTLDEGVLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGE
                     H"
ORIGIN

US-10-622-516-2 (1-478) x AK095492 (1-3195)

Alignment Scores:
Pred. No.:                  1.3e-195
Score:                      2403.50         Length:         3195
Percent Similarity:         89.81%          Matches:        467
Best Local Similarity:      89.81%          Conservative:   0
Query Match:                95.72%          Mismatches:     0
DB:                         9               Indels:         53
                                            Gaps:           2
```

```
Qy     1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
Db    68 ATGCCTCGGATCACTTCCTTGAATGCAGAAGCTTGCTGGCCAAAAGAGTGGGGAATTGTT  127

Qy    21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
Db   128 GCCCTTGAGATCTATTTCCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT  187

Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr   60
Db   188 GGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGTCAGGCCAAGATGGGCTTCTGCACA  247

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
Db   248 GATAGAGAAGATATTAACTCCCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT  307

Qy    81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys  100
Db   308 AACCTTTCCTATGATTGCATTGGGCGCCTGGAAGTTGGAACAGAGACAATCATCGACAAA  367

Qy   101 SerLysSerValLysAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle      120
Db   368 TCAAAGTCTGTGAAGAACTTAATTTGAAGACTCGTTTTGAAGAGCTCGGGAATACAGATATA 427

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal   140
Db   428 GAAGGAATCGACACAACT------------AATGCTGTT                        454

Qy   141 AsnTrpIleGluSerSerSerTrpAsp                                    149
Db   455 AACTGGATTGAGTCCAGCTCTTGGGATGGACCGGTATGCCCTGGTAGTTGCAGGAGATATT 514

Qy   149 ------                                                        149
Db   515 GCTTATATGCCACAGGAAAATGCTACAGGGGAGGTTGGAGCAGTAGCTCTGCTA        574

Qy   150 ------GlyLeuArgGlyThrHisMetGlnHis                              158
Db   575 ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT  634

Qy   159 AlaTyrAspPheLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu     178
Db   635 GCCTATGATTTTAAACCAGACATGCTATCTGAATATCCTATAGTAGATGGGAAACTC     694

Qy   179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle  198
Db   695 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC  754

Qy   199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe  218
Db   755 CATGCCCAGTGGCAGAAAGAAGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC  814

Qy   219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu  238
Db   815 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG  874

Qy   239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
Db   875 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC  934

Qy   259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet  278
Db   935 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAAAAGGCATTTATG  994

Qy   279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln  298
Db   995 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAAGCATCTTTACTTGTATCAAATCAA 1054

Qy   299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr  318
Db  1055 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1114

Qy   319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla  338
Db  1115 TCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTTCTGGGCTTGCT 1174

Qy   339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
Db  1175 GCCACTCTGTACTCTCTTAAAGTCACAACAAGATGCCACACCTGGGTCTGCTCTTGATAAA 1234

Qy   359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro  378
Db  1235 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA 1294

Qy   379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle  398
Db  1295 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAACTATATT 1354

Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu  418
Db  1355 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1414

Qy   419 LysHisArgArgThrTyrAlaArgProThrProAsnAspAspThrLeuAspGluGlyVal  438
```

RESULT 7
AX700129

LOCUS   AX700129   3275 bp   DNA   linear   PAT 03-APR-2003
DEFINITION   Sequence 15 from Patent EP1284298.
ACCESSION   AX700129
VERSION   AX700129.1 GI:29536069
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Rattus.
REFERENCE   1
  AUTHORS   Brooksbank,R.A., Dixon,A.K., Lee,K. and Pinnock,R.D.
  TITLE   Identification and use of molecules implicated in pain
  JOURNAL   Patent: EP 1284298-A 15 19-FEB-2003;
                WARNER-LAMBERT COMPANY (US)
FEATURES   Location/Qualifiers
     source   1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
                     /note="Cytosolic 3-hydroxy 3-methylglutaryl coenzyme A
                     synthase"
ORIGIN

US-10-622-516-2 (1-478) x AX700129 (1-3275)

Alignment Scores:
  Pred. No.:
  Score: 2362.00
  Percent Similarity: 89.62%
  Percent Identity: 87.50%
  Best Local Similarity: 94.07%
  Query Match: 6
  DB:

Db: AX700129 (1-3275)

Alignment Scores:
  Score: 2362.00

[Sequence alignment data]

```
Qy     459 ProArgLeuProAlaThrAlaAlaGluProGlnAlaAlaValIleSerAsnGlyGluHis 478
           ||||||||||||||||||||||||||||||:::
Db    1592 CCAAGACTTCCTGGGACCCGCAAACCTGAGTCGGGTTCATCAGTAACGGGGGAGCAC 1651
```

RESULT 8
AX827303

```
Db    1592 CCAAGACTTCCTGGGACCCGCAAACCTGAGTCGGGTTCATCAGTAACGGGGGAGCAC 1651
```

LOCUS   AX827303   3275 bp   DNA   linear   PAT 12-DEC-2003
DEFINITION   Sequence 37 from Patent EP1344834.
ACCESSION   AX827303
VERSION   AX827303.1 GI:39837392
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1
AUTHORS   Boess,F., Suter-Dick,L. and Wolf,D.
TITLE   Methods for the toxicity prediction of a compound
JOURNAL   Patent: EP 1344834-A 37 17-SEP-2003;
             F. HOFFMANN-LA ROCHE AG (CH)
FEATURES   Location/Qualifiers
     source   1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:10116"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 4.77e-192
Score: 2362.00
Percent Similarity: 89.62%
Best Local Similarity: 87.50%
Query Match: 94.07%
DB: 6

US-10-622-516-2 (1-478) x AX827303 (1-3275)

Length: 3275
Matches: 455
Conservative: 11
Mismatches: 12
Indels: 42
Gaps: 1
```

```
Qy       1 MetProGlySerLeuProLeuAsnAlaGlyAlaAlaCysTrpProLysAspValGlyIleVal  20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGCCTGGGTCGCTACCTCCTTTGAATGCAGAGGCTTGCTGGCCCAAAGATGTGGGAATCGTT 151

Qy      21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGlnLeuGlnLysTyrAsp  40
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     152 GCCCTTGAAATCTACTTTCCTTCTCAGTATGTTGATCAAGCTCAGTTGCAAAAATACGGAT 211

Qy      41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr  60
           ||||||||||||||||||||||||||:::||||||||||||||||
Db     212 GGTGTAGATGCTGGAAAGTATACCATTGGGCTGGGCCAGGCGATGGGCTTCTGCACG 271

Qy      61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGlyArgAsn  80
           ||||||||||||||||||||||||||||||:::||||||||||||||
Db     272 GATCGCGAAGACATCAACTCTCTGCCTGACTGTGGTTCAGAAACTGATGGGAGAGAAT 331

Qy      81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
           ||||||||:::||||||||||||||||||||||||||||||||||
Db     332 AGCCTTTCCTATGACTGCATTGGACGGCTGGAAGTCGGGACAGAGACAAATCATCGACAAA 391

Qy     101 SerLysSerSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
           ||||||||||:::|||||||||||||||||||||||||||||||||
Db     392 TCAAAATCCGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCTGGGAATACAGATATA 451

Qy     121 GluGlyIleAspThrThrAsnAlaCysTyrGlyThrAlaAlaValPheAsnAlaAlaVal 140
           |||||||||||||||||||||||||||||||||||||||||||||
Db     452 GAAGGAATAGATACAACTAATGCATGCTATGGGGCACAGCCCGCTAAGAAAGTG 511

Qy     141 AsnTrpIleGluSerSerSerTrpAsp------------- 149
           |||||||||||||||||||||||||
Db     512 AACTGGATCGAATCCAGCTCTTGGGATGGACCGATACGGCTTTGGTTAGTTGCACGAGACATC 571

Qy     149 ------------- 149
```

RESULT 9
RNCHMCA
LOCUS   RNCHMCA   3275 bp   mRNA   linear   ROD 18-JUN-2003

```
Db     572 GCTATATATGCCTCAGGAAACGCCAGGCTACAGGTGGGAGTTGGGACGTGTGGGCTGGCTA 631
           |||||||||||||||||||||||||||||||||||||||||||||
Qy     150 -----------------GlyLeuArgGlyThrHisMetGlnHis 158
                            ||||||||||||||||||
Db     632 ATTGGGGCCCAAAATGCTCCTGTGAATTTTTGACCGAGGGGCTTCGTTGGGACACACATGCAGCAT 691

Qy     159 AlaThrTyrAspPheTyrLysProAspMetLeuSerGluThrTyrProIleValAspGlyLysLeu 178
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db     692 GCCTACGGACTTTTACAAGCCTGACATGCTCTCTGAACATACCCCTGTTGGGATGGGAAAACTC 751

Qy     179 SerIleGlnCysTyrLeuSerArgAlaLeuAspArgCysTyrSerValTyrCysGlyLysIle 198
           ||||||||||||||||||||||||||||||||||||||||:||:|||||||
Db     752 TCCATACAGTGCTACCTCAGCCAGTTGGACCGCTGCTATTCCTGTCTACCGCAAAAGATC 811

Qy     199 HisAlaGlnIleTrpGlnLysGlyLysGlyValAsnAspAspPhePheThrLeuAsnAspPhePheGlyPhe 218
           ||||||||||||||||||||||||||||||||||||||||||||
Db     812 CGGGCCCAGTTGGCCAGAAAGGGAGGGAAAGGAATAAAAGATTTTACCCCGAATGATTTTGGCTTC 871

Qy     219 MetIlePheHisSerProTyrCysGlyLeuGluValGlnLysSerLeuValAlaArgMetLeuLeu 238
           |||||||||||:::|||||||||||||||||||||||||||||||||
Db     872 ATGATCTTTCACTCGCCATACTGTGGACTGGAAGTCCAAAAGTCTCTAGCCTGTCGATGCTGAATCAA 931

Qy     239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleLeuTyrSerGlyTyrLeuGlnTyr 258
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     932 AATGACTTTCTTAACGATCAAAACAGAGACAAAAACAGTAATTACAGTGGCTACCTGCAGTAC 991

Qy     259 PheGlyAspValLysLeuGluLeuGluLysAspThrTyrPheAspProArgAspValGlyLysAlaPheMet 278
           ||||||||||||||:::||||||||||||||||||||||||||||||
Db     992 TTTGGGGATGTGAAATTAGAAGATACTTACTTCGACCAGAGAGATGTGGAAAAGGCATTTATG 1051

Qy     279 LysAlaSerSerGluLeuPheSerGlyLysSerThrLysAlaSerLeuLeuValSerAsnGln 298
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1052 AAGGCTAGTGCTGAGCTATTCAACCAGAAAAGCACATCTTTGCTGTTCTGTCTATCGAATCAA 1111

Qy     299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
           ||||||||||||||||||||||||||||||||||||||||||||
Db    1112 AATGGAAACATGTACACATCCTCTGTATACGGTTCCCTTGCTTCTGTTCTGGCCACGTAC 1171

Qy     319 SerProGlnLeuLeuAlaGlyLysAspValPheSerThrTyrGlySerGlyLeuVal 338
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1172 TCACCTCAACAGTTGGCCGGGAAGGATGTTTCTACGGTTCGGTTCTGGCTTGGCT 1231

Qy     339 AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 GCCACACTCTACTCCCTTAAAGTCACAGATGCCACAAGGAATCGTGCTCTTGACAAA 1291

Qy     359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
           ||||||||||||||||||||||||||||||||||||||||||||
Db    1292 ATAACAGCAAGTTTATGTGACCTTAAGTCAAGGCTTGACTCAAGAACGTGTGGCACCA 1351

Qy     379 AspValPheAlaGluAsnMetLysLeuValArgGluLeuAspThrHisHisLeuValAsnTyrIle 398
           |||||||||||||||||||||||||||||||||||||||||||||
Db    1352 GACGTCTTTGCTGAAAACATGAAGCTGGTTAGAGAGCTGGACACACATCACTTAGCCAACTATATT 1411

Qy     399 ProGlnGlySerIleAspSerLeuPheGluGlyTyrTrpTyrLeuValArgValAspGlu 418
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 CCCCAGTGTTCAATAGATTCACTCTTCGAAGGATACTGGTATCTAGTCAGAGTGGATGAA 1471

Qy     419 LysHisIleArgArgThrTyrAlaArgProThrProAsnAspAspThrLeuAsnGluGly 438
           ||||||||||||||||||||||||||:::|||||||||||||||||
Db    1472 AAGCACAGAAGAACTTACGCCCGGCCTCCCAACAAATGACCACACAGTTTGGATGAAGGA 1531

Qy     439 ValGlyLeuValHisSerAsnIleIleProSerProAlaLysLysVal 458
           ||||||||||||||||||||||||||||||||||:::
Db    1532 GTGGGACTTGTGCATTCAAACAACATCATCCCAAGCCCGGCTAAGAAAGTG 1591

Qy     459 ProArgLeuProAlaThrAlaAlaGluProGlnAlaAlaValIleSerAsnGlyGluHis 478
           |||||||||||||||||||||||||:::
Db    1592 CCAAGACTTCCTGGGACCCGCAAACCTGAGTCGGGTCGTCATCAGTAACGGGGGAGCAC 1651
```

```
DEFINITION   Rattus norvegicus mRNA for cytosolic 3-hydroxy 3-methylglutaryl
             coenzyme A synthase (EC 4.1.3.5).
ACCESSION    X52625
VERSION      X52625.1  GI:55946
KEYWORDS     cytosolic 3-hydroxy 3-methylglutaryl coenzyme A synthase.
SOURCE       Rattus norvegicus (Norway rat)
  ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE    1
  AUTHORS    Ayte,J., Gil-Gomez,G. and Hegardt,F.G.
  TITLE      Nucleotide sequence of a rat liver cDNA encoding the cytosolic
             3-hydroxy-3-methylglutaryl coenzyme A synthase
  JOURNAL    Nucleic Acids Res. 18 (12), 3642 (1990)
   MEDLINE   90301491
    PUBMED   1972979
  REMARK     (b 1-38,40,41-62,64-3275)
REFERENCE    2
  AUTHORS    Hegardt,F.G.
  TITLE      Direct Submission
  JOURNAL    Submitted (17-APR-1990) Hegardt F.G., University of Barcelona, Unit
             of Biochemistry, School of Pharmacy, Placa Pius XII, s/n. 08028
             barcelona, Spain
  REMARK     (b 1-38,40,41-62,64-3275)
REFERENCE    3  (bases 1 to 3275)
  AUTHORS    Haegardt,F.G.
  TITLE      Direct Submission
  JOURNAL    Submitted (30-JUL-1990)
  REMARK     revised by [3]
FEATURES             Location/Qualifiers
     source          1..3275
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /clone="lambda-cCS1"
                     /tissue_type="liver"
     old_sequence    31..70
                     /note="revised by [3]"
                     /citation=[3]
     CDS             92..1654
                     /EC_number="4.1.3.5"
                     /codon_start=1
                     /product="cytosolic 3-hydroxy 3-methylglutaryl coenzyme A
                     synthase"
                     /protein_id="CAA36852.1"
                     /db_xref="GI:55947"
                     /db_xref="GOA:P17425"
                     /db_xref="SWISS-PROT:P17425"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELKYDGVDA
                     GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDI
                     AIYASGNARPTGGVGAVALLIGPNAPVIFDRGLRGTHMQHAYDVFKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASAELFNQKTKAS
                     LLVSNQNGNMYTSSVVGSLASVLAQYSPQQLAGKRIGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITALSCLDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPSTNDHSLDEGVLVHSNTATEHIPSPAKKVPRLPATSG
                     EPESAVISNGEH"
ORIGIN
```

```
US-10-622-516-2 (1-478) x RNCHMCA (1-3275)

Alignment Scores:
Pred. No.:                          4.77e-192
Score:                              2362.00        Length:       3275
Percent Similarity:                 89.62%         Matches:      455
Best Local Similarity:              87.50%         Conservative: 11
Query Match:                        94.07%         Mismatches:   12
DB:                                 10             Indels:       42
                                                   Gaps:         1
```

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     92 ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCTTGCTGGCCAAAAGATGTGGGAATCGTT 151

Qy     21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuLysTyrAsp 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    152 GCCCTTGAAATCTACTTTCCTTCTCAGTATGTTGATCAAGCTGAGTTGAAAAATACGAT 211

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaArgMetGlyPheCysThr 60
          |||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db    212 GGTGTAGATGCTGGGAAAGTATACCATTGGCCTGGGCCAGGCGACAATGGGTTCTGCACG 271

Qy     61 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 80
          ||||:||||||||||||:||||::|||||||||||||||||||||||||:||||||||
Db    272 TCAAAAATCCGTGAAGTCTAAATTTGATGCAGGCAGTTTGAGGAGTCTGGAATACAGATAT 331

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
          ::::|||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db    332 AGCCTTTCCTATGACTGCATTGGGCGGCTGGAAGTGGGCACAGAGACAATCATCGACAA 391

Qy    101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
          |||||:|||||||||||:||||::|||||||||||||||||||||||||:||||||||
Db    392 TCAAAAATCCGTGAAGTCTAAATTTGATGCAGGCAGTTTGAGGAGTCTGGAATACAGATAT 451

Qy    121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
          |||||||||||||||||||:|||||:||||||||||||||||||||||||||:|||||||
Db    452 GAAGGAATAGATACAACTAATGCATGTTATGGGGGAACAGCAGCCGTCTTCAATGCTGTG 511

Qy    141 AsnTrpIleGluSerSerSerTrpAsp--------------------------- 149
          |||||||||||||||||||||||||||
Db    512 AACTGGATCGAATCCAGCTCTTGGGATGGACGATACGCTTGGGTAGTTGCAGGAGACATC 571

Qy    149 -------------------------------------------------- 149
Db    572 GCTATATATGCCTCAGGAAACGCCAGGCCTACAGGTGGAGTTGGAGCTGTGGCCCTGCTA 631

Qy    150 -----------------------------GlyValLeuArgGlyThrHisMetGlnHis 158
                                       |||||||||||||||||||||||||||||||
Db    632 ATTGGGCCAAATGCTCCTGTAATTTTTGACCGAGGGCTTCGTGGGACACACATGCAGCAT 691

Qy    159 AlaTyrAspValPheLysProAspMetLeuSerGluTyrProValValAspGlyLysLeu 178
          |||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db    692 GCCTACGACGTTTTCAAGCCTGACATGCTCTCTGAATATCCCGTTGGTAGATGGGAAACTC 751

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          ||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db    752 TCCATACAGTGCTACCTCAGCGCATTGGACCGCTGTTATTCTGTCTATTGTAAAAGAATC 811

Qy    199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
          |||||||||||||||||||||:::|||||:|||||||||||||||||||||||||||||
Db    812 CGGGCGCCAGTGGCAGAAAGGGAAGGATAAAGATAAAGATTTTACCCTGAATGATTTTGGCTTC 871

Qy    219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    872 ATGATCTTTCACTCGCCATACTGTAAACTGGTGCAGAAATCTCTAGCTAGGATGTTCCTG 931

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
          ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db    932 AATGACTTTCTTTAACCGATCAAAAACAGAGACAATGAAAGTGGGCCGCTGGAAGCC 991

Qy    259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992 TTTGGGGATGTGAAATTAGAAGATACTTACTTCGACAGAGATGTGGAAAAGGCATTTATG 1051

Qy    279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
          ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db   1052 AAGGCTAGTGCTGAGCTATTCAACCAGAAAACAAAGGCATCTTTGCTTGTATGCAATCAA 1111

Qy    299 AsnGlyAsnMetTyrThrSerSerValValGlySerLeuAlaSerValLeuAlaGlnTyr 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1112 AATGGAAACATGTACACATCTTCTGTATCCGTTCCCTTGCTTCTGTTTCTGGCACAGTAC 1171

Qy    319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1172 TCACCTCAACAGTTGGCCGGAAGAGGATTGGAGTTCTCTTACGGTTCTGGCTTGGCT 1231
```

RESULT 10

```
LOCUS       BC023851                2703 bp    mRNA    linear   ROD 06-OCT-2003
DEFINITION  Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36620 IMAGE:5347038), complete cds.
ACCESSION   BC023851
VERSION     BC023851.1  GI:23271493
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2703)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE   22388257
   PUBMED    12477932
REFERENCE   2  (bases 1 to 2703)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Jeffrey Green M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stanripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 55 Row: h Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21122370.
FEATURES             Location/Qualifiers
     source          1..2703
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:36620 IMAGE:5347038"
                     /tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
                     ductal carcinoma. 5 month old virgin mouse."
                     /clone_lib="NCI_CGAP_Mam6"
                     /lab_host="DH10B"
                     /vector: pCMV-SPORT6"
     gene            1..2703
                     /gene="Hmgcs1"
                     /note="synonym: MGC156662"
                     /db_xref="Locus:ID:208715"
                     /db_xref="MGI:107592"
     CDS             149..1711
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
                     1"
                     /protein_id="AAH23851.1"
                     /db_xref="GI:23271494"
                     /db_xref="LocusID:208715"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                     GKYTIGLGQARMGFCDREDINSLCLTVVQKLMERHSLSYDIGIRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                     AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLLEDTYFDRDVEKAPMKASSELFNQKTKAS
                     LLVSNQNGNMYTSSVYGSLASVLAQYSPQGLAGKRVGVFSYGSGLAATLYSLKVTODA
                     TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRTYARPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                     ESSAVISNGEH"
     misc_feature    185..1555
                     /note="HMG_CoA_synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CDD:pfam01154"
ORIGIN
```

```
Qy   339  AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrrProGlySerAlaLeuAspGlyLys  358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1232  GCCACAACTCTACTCCCTTAAAGTCACACAAGATGCCACCACCAGGATCTGCTCTTGACAAA  1291

Qy   359  IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro  378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1292  ATAACAGCAAGTTTATGTGACCTTAAGTCAAGGCTTGACTCAAGGAACGGTGTGTGGCACCA  1351

Qy   379  AspValPheAlaGluAsnMetLeuArgGluArgHisHisLeuValAsnTyrIle  398
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1352  GACGTCTTTGCTGAAAACATGAAGCTCGAGAGGAGGCACACAATCACTTAGCCAACTATATT  1411

Qy   399  ProGlnGlySerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu  418
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1412  CCCCAGTGTTCAAATAGATTCACTCCTTCGAAGGAACATGGTATCTAGTCAGAGTGGATGAA  1471

Qy   419  LysHisSerArgArgThrTyrAlaArgArgProThrProAsnAspAlaAspThrLeuAspGlyGly  438
          |::||||||||||||||||||||||||||||||||::||||||||||||||||
Db  1472  AAGCACAGAAGAACTTACGCCCGGCGTCCCTCCACAAATGACCCACGTTTGGATGAAGGA  1531

Qy   439  ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal  458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1532  GTGGGACTTGTGCAATCAAACACAGCTACAGAGCATATTCCAAGCCCCGGCTAAGAAAGTG  1591

Qy   459  ProArgLeuProAlaThrAlaAlaGluProGluLeuAlaValIleIleSerAsnGlyHis  478
          |||||||||||::||:::||||||||||||||||||||||||||||||||
Db  1592  CCAAGACTTCCTGCCGACCTCGGCTGTCATCAGTAACGGGGAGGAGCAC  1651
```

Alignment Scores:
Pred. No.:
Pred: 5.8e-191   Length: 2703
Score: 2348.00   Matches: 452
Percent Similarity: 89.42%   Conservative: 13
Percent Identity: 86.22%   Mismatches: 13
Best Local Similarity: 93.51%   Indels: 42
Query Match: 10   Gaps: 1
DB:

US-10-622-516-2 (1-478) x BC023851 (1-2703)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluThrAlaCysTrpProLysAspValGlyIleVal 20
          ATGCCTGGGTCACTTCCTTTGAATGCAGAGACGGCTTGCTGGCCCAAAGATGTGGGAATCGTT
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  149 ...                                                              208

Qy   21 AlaLeuGlyIleThrTyrProSerGlnThrTyrValAspAspLeuGluGlyLysTyrAsp    40
          GCCCTGGAAATCACTTACCCTTCTCAAATATGTCGGATGAAGGCCAAAAATACGAT
          :::::::::::::::::::::::
Db  209 ...                                                              268

Qy   41 GlyValAlaProAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
          GGTGTAGCACCTGCAGGAAAGTATACCATCGGCCTGGGCCAGGCCAAGATGGGCTTCTGCACG
          ::::::::::::::::::::::::::::::::::::::::
Db  269 ...                                                              328

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn    80
          GATCGTGAAGACATCAACTCTCTTTGCCTGACTGTGGTTCAGAACTGATGGAGAGACAT
          ::::::::::::::::::::::::::::::::
Db  329 ...                                                              388

Qy   81 AsnLeuSerThrTyrAspCysIleGlyPheArgLeuGluValGlyGlyGlyGlnGlnAsn    100
          AACCTTTCCTATGATTGCATTGGGTTCCGGCTGGAAGTTGGAGGAGGTCAGCAAAATACGAT
          :::::::::::::::::::::::::::
Db  389 ...                                                              448

Qy  101 GluGluIleAspAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValValPheAsnAlaVal 120
          GAAGGAAATAGATGATACAACAAATGCATGCTATGGGGGCACAGCTGCAGTCGTCTTCAATGCCGTG
          :::::::::::::::::::::::::::::::
Db  449 ...                                                              508

Qy  121 SerLysSerValGlyIleGlyArgCysIleGlyValLeuGlyGlyGlyValGlnThrGluLysAsnThrAsp 140
          TCGAAATCAGTGGAAGTTCTAATTTGGAGTGGTGGGGGCGGGGGTCAGACTGAAGATAACAGATATA
          ::::::::::::::::::::::::
Db  509 ...                                                              568

Qy  141 AsnTrpIleGluSerSerSerTrpAsp                                       149
          AACTGGGGTCGAATCCAGCTTCTTGGGGATGGACGATATGCTTCTCAGGAGACATT
          ::::::::::::::
Db  569 ...                                                              628

Qy  149 ------------------------------------------------------------ 149
          
Db  629 GCTATATATGCCACAGGAAATGCCAGACCTACAGGTGGAGCTGGAGCTGGCCCTGCTA    688

Qy  150 ------------GlyLeuArgGlyThrThrHisMetGlnHis                        158
          ------------GGTCTCCGTGGGACACACATGCAGCAT
          |||
Db  689 ...                                                              748

Qy  159 AlaTyrAspMetLeuSerGluTyrProProValProProLysAspIleGlyIleLeu         178
          GCCTATGACATGCTTTCAGAGTACCCGGTTCCACCGAAAGGATATTGGGATGGACAGCTTC
          :::::::::::::::::::::::::::
Db  749 ...                                                              808

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle     198
          TCCATACAGTGCTACCTCAGCGCCCTGGACCGGCTATTCGTGTACCGCAAAAGATC
          :::::::::::::::::::::::::::
Db  809 ...                                                              868

Qy  199 HisAlaGlnTrpGlnLysGluGlyAsnAspAlaProPheThrLeuAsnAspAspPheGlyPhe 218
          CGTGCCCAGTGGCAGAAAGGAGGAAATGATGCACCATTTACCCTGAATGATGATTTTGGCTTC
          :::::::::::::::::::::::::::::::::::::::
Db  869 ...                                                              928

Qy  219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu      238
          ATGATCTTCCACTCACCATATTGTAAACTGGTGCAGAAATCTCTAGCTCGGATGTTCCTG
          :::
Db  929 ...                                                              988

Qy  239 AsnAspPheLeuAsnAsnGlnAsnArgAspProLysAsnSerIleTyrSerGlyLeuGluAla 258
          AATGACTTTCTTAATGATCAAAACAGAGATCCAAAAAACAGTATTTACAGTGGACTCGAAGCC
          :::::::::::::::::::::::::::::::::::::::::
Db  989 ...                                                              1048

Qy  259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysLysAlaPheMet 278
          TTTGGGGACGTTAAATTGGAAGATACTTACTTTGACAGAGATGTAGGAAAGAAAGCATTTATG
          :::::::::::::::::::::::::::::::
Db  1049 ...                                                             1108

Qy  279 LysAlaSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuLeuSerAsnGln         298
          AAGGCTAGTTCTGAGCTATTCAGCCAGAAAACAAAAGCGTCTTTGCTTGTGTCTAAATCG
          ::::::::::::::::::::::::::::::
Db  1109 ...                                                             1168

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr     318
          AATGGAAATATGTACACATCCTCTGTCTATGGGTCCCTGGCTTCTGTCCTGGCACAGTAC
          :::::::::::::::::::::::::::::::::
Db  1169 ...                                                             1228
```

```
Qy  319 SerProGlnIleLeuAlaGlyLysArgIleGluGlyValPheSerTyrGlySerGlyLeuAla 338
          TCACCTCAGCAGTTGGCAGGTAAGCGGATTGAAGGGGTGTTCTCTTACGGTTCAGGGCTTGGCT
          :::::::::::::::::::::::::::::::::::::::
Db 1229 ...                                                             1288

Qy  339 AlaThrLeuTyrSerLeuTyrLeuLysLysValThrThrProGlySerAlaLeuAspLys     358
          GCCACGCTGTACTCCCTTAAAGTCACACAAGACGGCGTTGCCCTTGATAAA
          ::::::::::::::::::::
Db 1289 ...                                                             1348

Qy  359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaAlaPro 378
          ATAACAGCAAGTTTATGTGACCTTAAAATCACGGCTGGAACTGTGTGGCACCG
          |||
Db 1349 ...                                                             1408

Qy  379 AspValPheAlaGluAsnMetLysLeuArgGluGluAspThrHisHisSerLeuValAsnTyrIle 398
          GATGTGTTTGCTGAAAACATGAAAGCTCAGAGAAGAAGGACACAACATCACCACTATATTATT
          :::::::::::::::::::::::::
Db 1409 ...                                                             1468

Qy  399 ProGlnGlySerIleAspSerSerLeuProGluGluProGlySerProGlySerAlaLeuGlu 418
          CCCCAGTGTTCAATAGACTCCACTCTTCCAGAGGAACCAGGAAGCCCAGGGAGCGCACTTGAA
          :::::::::::::::::::::::::::::::::::::::
Db 1469 ...                                                             1528

Qy  419 LysHisArgAspThrArgTyrArgAlaArgProThrProAsnAspAspThrHisHisSerLeuValAsnTyr 438
          AAACAGCAGAGAGACTTACGCCGCCGGCCCACTCCTAATGATGACAGTTGGTGATGAAGGA
          ::::::::::::::::::::
Db 1529 ...                                                             1588

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal    458
          GTGGGGCTCGTGCATAGTAACACAGCAGAGCATATTCCAAGCCCTGCTAAGAAAGTG
          :::::::::::::::::::::::::::
Db 1589 ...                                                             1648

Qy  459 ProArgGluLeuProAlaThrAlaAlaGluProGluAlaValIleIleSerAsnAsnGlyValHis 478
          CCAAGAGAACTCCCTGCAACCGCCGCCGAATCTGAAGCAGTCATCATCAGTAACGGGGGAGCAC
          ::::::::::::::::::::::::::::::::
Db 1649 CCAGAGACTCCCTGCAACCTCGGCCCCAAATTCCGGGAATCAGCTCGTGGAGGAGCAC     1708
```

REMARK
　　Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
　　USA
　　NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT
　　Contact: MGC help desk
　　Email: cgapbs-r@mail.nih.gov
　　Tissue Procurement: The Cepko Laboratory
　　cDNA Library Preparation: Life Technologies, Inc.
　　cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
　　DNA Sequencing by: Baylor College of Medicine Human Genome
　　Sequencing Center
　　Center code: BCM-HGSC
　　Web site: http://www.hgsc.bcm.tmc.edu/cdna/
　　Contact: amg@bcm.tmc.edu
　　Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
　　Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
　　A.N., Gibbs, R.A.
　　Clone distribution: MGC clone distribution information can be found
　　through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
　　Series: IRAK Plate: 53 Row: k Column: 24
　　This clone was selected for full length sequencing because it
　　passed the following selection criteria: matched mRNA gi: 22122370.

FEATURES　　　　Location/Qualifiers
source　　　　　1..3202
　　　　　　　　/organism="Mus musculus"
　　　　　　　　/mol_type="mRNA"
　　　　　　　　/db_xref="taxon:10090"
　　　　　　　　/clone="MGC:36662 IMAGE:5367786"
　　　　　　　　/tissue_type="Eye, retina, mouse strain C57Bl\6"
　　　　　　　　/clone_lib="NIH_MGC_94"
　　　　　　　　/lab_host="DH10B"
　　　　　　　　/note="Vector: pCMV-SPORT6"
gene　　　　　　1..3202
　　　　　　　　/gene="Hmgcs1"
　　　　　　　　/note="synonym: MGC36662"
　　　　　　　　/db_xref="GI:21706866"
　　　　　　　　/db_xref="LocusID:208715"
　　　　　　　　/db_xref="MGI:107592"
CDS　　　　　　　25..1587
　　　　　　　　/codon_start=1
　　　　　　　　/product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
　　　　　　　　1"
　　　　　　　　/protein_id="AAH34317.1"
　　　　　　　　/db_xref="GI:21706866"
　　　　　　　　/db_xref="LocusID:208715"
　　　　　　　　/translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKIDGVDA
　　　　　　　　GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
　　　　　　　　SVKSNLMQLFEESGNTDIGITDTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
　　　　　　　　AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDYYKPDMLSEYPVVDG
　　　　　　　　KLSIQCYLSALDRCYEVYRKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSIA
　　　　　　　　RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFNQKTKAS
　　　　　　　　LLVSNQNGMYTSSVYGSLASVLAQYSPQQLAGKRVGVFSYGSGLAATLYSLKVTQDA
　　　　　　　　TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
　　　　　　　　GTWYLVRVDEKHRRTYARRPTNDHSLDEGMLVHSNTATEHIPSPAKVPRLPATSA
　　　　　　　　ESESAVISNGEH"
misc_feature　61..1431
　　　　　　　　/note="HMG_CoA_synt; Region:
　　　　　　　　Hydroxymethylglutaryl-coenzyme A synthase"
　　　　　　　　/db_xref="CDD:pfam01154"

ORIGIN

Alignment Scores:
Pred. No.:
Score:　　　　　　　　2348.00
Percent Similarity:　　89.42%
Best Local Similarity: 86.92%
Query Match:　　　　　　93.51%
DB:

Length:　　　　3202
Matches:　　　　452
Conservative: 13
Mismatches:　　13
Indels:　　　　42
Gaps:　　　　　1

7.29e-191

US-10-622-516-2

Qy　　1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
　　　　||||||||||||||||||||||||||||||||:::||||||||:||||||||||||||||
Db　　25 ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCTTGCTGGCCAAAAGATGTGGGAATCGTT 84

Qy　　21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　85 GCCCTTGAAATCTACTTTCCTTCTCAATATGTCGACCAAGCTGAGTTGGAAAATTACGAT 144

Qy　　41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaArgMetGlyPheCysThr 60
　　　　||||||||||||:||||||||||||||||||||||:|||||||||||||||||||||:||
Db　　145 GGTGTAGATGCTGGGAAAGTTATACCATCGGGTCTGGGGCAAGCCAGGATGGGCTTCTGCACG 204

Qy　　61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
　　　　|||||||||||||||||||||||:||:||||||||||||||||||||||||||:::||:::
Db　　205 GATCGTGAAGACATCAACTCTCTTTGCCTGACTGTGGTTCAGAAACTGATGGAGAGAAAT 264

Qy　　81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
　　　　:::||||||||||||||||:|||||||||||||||||||||||:|||||||||||:
Db　　265 AGCCTTTCCTATGATTGCATTGGGCGGCTAGAAGTTGGAACAGAGACAATCATCGACAAA 324

Qy　　101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
　　　　||||||||||||||||||||||||:||||||||:||||||||||||||||||||||||||
Db　　325 TCGAAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCTGGGAATACAGATATA 384

Qy　　121 GlyIleThrAspThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
　　　　||||||||||||||||||||||||||||:||||||||||:||||||||||:|||||||
Db　　385 GGAAGAAATACAGATACAAACGCATGCTATGGGGGCACAGCTGCAGTCTTCAATGCCGTG 444

Qy　　141 AsnTrpIleGluSerSerSerTrpAsp-------------------------- 149
　　　　|||||:|:::||||||:|||||||||:
Db　　445 AACTGGGTCGAATCCAGCTCTTGGGATGGGAGGTACGCTCTTGTTGTCGCAGGAGACATT 504

Qy　　150 ----------------------GlyLeuArgGlyThrHisMetGlnHis 158
　　　　　　　　　　　　　　　　　|||||||||||||||:||
Db　　505 GCTATATATGCCACAGGAAATGCCAGACCTACAGGTGGAGTTGGAGCTGTGGCCCTGCTA 564

Qy　　159 AlaTyrAspTyrTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
　　　　||||||||||||||||||||||:|||||||||||||||:||||||||||||||||||
Db　　565 GCCTATGACTATTACAAGCCTGACATGCTCTCCGAGTACCCTGTGGTTGACGGAAAGCTC 624

Qy　　179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrGluValTyrArgLysIleIle 198
　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db　　625 TCCATACAGTGCTATCTCTCAGCGCTGGACCGGTGCTATTCGTGTCTACCGGAAAAGATC 684

Qy　　199 HisAlaGlnTrpGlnLysGluGlyLysAspAspAspPheThrLeuAsnAspPheGlyPhe 218
　　　　||||||||||:|||||||||||||||||||:|||||||||||||||||||||||||
Db　　685 CGTGCCCAGTGGCAGAAAGAGGGAAAGGACAAAGGATTTTACCCTGAATGAATTTTGGCTTC 744

Qy　　219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerIleAlaArgMetLeuLeu 238
　　　　||||||||||||||:||||||||||:||||||||||||||||||||||||||||||
Db　　745 ATGATCTTTCACTCACCATATTGTAAACTGGTGCAGAAATCTATAGCTCGGATGTTCCTG 804

Qy　　239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
　　　　||||||||||||||||:|||:||||||:|||||||||||||||||||:||:
Db　　805 AATGACTTTCTTAATGATCAAAACAGAGATAAAAACAGTATTTACAGTGGACTGGAAGCC 864

Qy　　259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
　　　　||||||||||||||||||||:||||:|:|||||||||||
Db　　865 TTTGGGGACGTTAAATTAGAGGATACTTACTTTGACAGAGATGTAGAAAAGGCATTTATG 924

Qy　　279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
　　　　||||||||||||||||:||||||||||||||||||||||||:||:||
Db　　925 AAGGCTAGTTCTGAGCTATTCAAACAGAAAACAAAAGCGTCTCTTGCTGTGTCTAATCAG 984

Qy　　299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
　　　　||||||||||||||||||||||||||||||||||||:|||||:
Db　　985 AATGGAAATATGTACACATCCTCGTCTTACGGTTCCCTGGCTTCTGTCCTGGCACAGTAC 1044

Qy　　319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
　　　　||||||||||||||||:||||||:||||||||||||||||||||||:
Db　　1045 TCACCTCAGCAGTTGGCAGGGAAGAGAATTGGAGTGTTCTCTTACGGTTCTGGGCTGGCT 1104

Qy
Db　　1105 TCACCTCAGCAGTTGGCAGGGAAGAGAGAATTGGAGTGTTCTCTTACGGTTCTGGGCTGGCT 1164

```
Qy    339  AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165  GCCACGCTGTACTCCCTTAAAGTCACAAGAGGCCCACACCAGGATCTGCCCTTGATAAA 1224

Qy    359  IleThrAlaSerLeuCysSerArgLeuLysSerArgGlnLeuAspSerArgThrGlyValAlaPro  378
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225  ATAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTCGACTCCAGAACTTGTGTGGCACCG 1284

Qy    379  AspValPheAlaGluAsnMetLysArgGluArgThrHisHisLeuValAsnTyrIle  398
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   1285  GATGTCTTTGCTGAAAACATGAAGCTCAGAGAGCGAACACACCTTAGCCAACTATATT 1344

Qy    399  ProGlnGlySerIleAspSerLeuPheGluGlyThrThrTyrLeuValArgValAspGlu  418
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345  CCCCAGTTGTTCAATAGACTCACTCTTTGAAGGAACGTGTATCTGGTGCGTGGATGAA 1404

Qy    419  LysHisHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly  438
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1405  AAACACAGAAGGACTTACGCCCGGCGGCCCCGGACCCTTCACAAATGACCACAGTTTGGATGAAGGA 1464

Qy    459  ProArgLeuProAlaThrAlaAlaGluProGlnAlaAlaValIleSerAsnGlyGluHis  478
           ||||||::::::|||||||||||||::::|||||||:::||||||::::||
Db   1525  CCAAGACTCCCTGCAAACCTCGGCCGGAATCTGAATCAGCTGCTCATCAGTAACGGGGAGCAC 1584

Qy    439  ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysHisValGlyVal  458
           ::::||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465  ATGGGGCTCGTGCATAGTAACACAGCATAATTCCAAGCCCTGCTAAAGCCTAAGAAAGTG 1524
```

RESULT 12
BC029693

```
LOCUS       BC029693              3260 bp    mRNA    linear   ROD 07-OCT-2003
DEFINITION  Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36525 IMAGE:5375374), complete cds.
ACCESSION   BC029693
VERSION     BC029693.1  GI:20988708
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3260)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 3260)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 53 Row: p Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 22122370.
FEATURES             Location/Qualifiers
     source          1..3260
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="MGC:36525 IMAGE:5375374"
                     /tissue_type="Mammary tumor." MMTV-LTR/INT3 model. 5 month
                     old mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..3260
                     /gene="Hmgcs1"
                     /note="synonym: MGC36662"
                     /db_xref="LocusID:208715"
                     /db_xref="MGI:107592"
     CDS             79..141
                     /codon_start=1
                     /product="Hmgcs1 protein"
                     /protein_id="AAH22693.1"
                     /db_xref="GI:20988709"
                     /db_xref="LocusID:208715"
                     /translation="MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKVDGVDA
                     GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHLSYDCIGRLEVGTETIIDKSK
                     SVKSNLMQLFESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                     AIYATGNARPTGVGAVALLIGPNAPLIFDRGLRGTHMQHAIPDFYKPDMLSEXPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWDKEGKQKDCFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDLNDQNRDKNSIYSGLEAFGDVKLEDYIFDRDVEKARFMKASSELFNQKTKAS
                     LIVSNQNGNMYTSSVYGGLASVLAQYSPQQLAGKRVGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITTASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPFINDHSILDEGMGLVHSNTATEHIPSPAKKVRLPATSA
                     ESESAVISNGEH"
     misc_feature    115..1485
                     /note="HMG_CoA_synt; Region:
                     Hydroxymethylglutaryl-coenzyme A synthase"
                     /db_xref="CDD:pfam01154"
ORIGIN
```

US-10-622-516-2 (1-478) x BC029693 (1-3260)

```
Alignment Scores:
Pred. No.:
Score:              7.47e-191
Percent Similarity:     2348.00        Length:            3260
Best Local Similarity:  89.42%         Matches:            452
Query Match:            86.92%         Conservative:        13
DB:                     93.51%         Mismatches:          13
                        10             Indels:              42
                                       Gaps:                 1
```

```
Qy      1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleValIleVal  20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     79  ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCTTGCTGGCCCAAAAGATGTGGGAATCGTT 138

Qy     21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Qy   139 GCCCTTGAAATCTACTTTCCTTCTCAAATATGTCGACTCAACGTGAGTTGGAAAATACCAT 198
Db   139 GCCCTTGAAATCTACTTTCCTTCTCAAATATGTCGACTCAACGTGAGTTGGAAAATACCAT 198

Qy   41 GlyValAspAlaGlyAlaGlyLysTyrThrIleGlyLysMetGlyPheCysThr 60
Db   41 GGTGTAGGATCCTGGAAAGTATACCATGGGCTGGGCGTCCAGCCGATGGGCTTCTGCACG 198

Db   199 GGTGTAGGATCCTGGAAAGTATACCATGGGCTGGGCGTCCAGCCGATGGGCTTCTGCACG 258

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db   259 GATCGTGAAGACATCAACTCTCTTTGCCTGACTGTGGTTCAGAAACTGATGGAGAGACAT 318

Qy   81 AsnLeuSerTyrArgAspCysIleGluGlyTyrArgLeuGluValGlyThrGlnThrIleIleIleGluAspLys 100
Db   319 AGCCTTTCCTATGATTGCATTGGGGGCGCCTAGAAGTTGGAACAGAATTCATCGACAAA 378

Qy   101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluLysGlyAsnThrAspIle 120
Db   379 TCGAAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCCGGGAATACAGAATATA 438

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal 140
Db   439 GAAGGAAATAGATACACAAACAAATGCATGTGATGGGGCCACACCGCTCTTCAATGCCGTG 498

Qy   141 AsnTrpIleGluSerSerSerTrpAsp----------------------- 149
Db   499 AACTGGGTCGAATCCAGCTCTTGGGATGGCAATAGTCGTCTGGTAGTTGCAGGAGACATT 558

Db   559 GCTATATATGCCCACAGGAAAATGCCAGACCTACAGTGTGGGCGGGCCCTGCTA 618

Qy   150 -----------GlyLeuArgGlyThrHisMetGlnHis 158
Db   619 ATTGGGCCAAAACGCTCCTCTAATTTTTGACCGAGGCTCCGTGGGAACACACATGCAGCAT 678

Qy   159 AlaThrTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
Db   679 GCCACATGACTTTTACAAGCCTGACATGCTCTCCGAGTACCCGATTGTGGATGGGAAAGCTC 738

Qy   179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
Db   739 TCCATACAGTGCTACCTCAGCGCCCTGGACCGCTGCTATTCTGTCTACCGAAAAAGATC 798

Qy   199 HisAlaGlnIleTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
Db   799 CGTGCCCAGTGGCAGAAAGAGGGAAACGATAAAGATTTTACCCTGAATGATTTTGGCTTC 858

Qy   219 MetIlePheHisSerProThrTyrCysLysLysGlnLysSerLeuLeuArgMetLeuLeu 238
Db   859 ATGATCTTTCACTCACCATATTGTAAAACTGGTGCACAAAATCTCAGCCTCGGATGTTCCTG 918

Qy   239 AsnAspPheLeuAsnAspGlnAsnArgAspSerLysAsnSerIleTyrSerGlyLeuGluAla 258
Db   919 AATGACTTTCTTAATGACCAGAACAGAGATAGTAAAAACAGTATTTACAGTGGACTGGAAGCC 978

Qy   259 PheGlyAspValLysLeuAspThrTyrPheAspAspGlyValGluLysAlaAlaPheMet 278
Db   979 TTTGGGGACGTTAAATTAGAAGATACTTACTTTGACGAGATGTTGAGAAAAGGCCATTTATG 1038

Qy   279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db   1039 AAGGCTAGTTCTGAGCTATTCAACCAGAAAACAAAGGCGTCTTTGCTTGTGTCTAATCAG 1098

Qy   299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db   1099 AATGGAAATATGTACACAAGCTCTGTATGGTTCCCTGGCTTCTGTTCCTGGCACAGTAC 1158

Qy   319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
Db   1159 TCACCTCACCAGTTGGCAGGGAAGAGGTTGGAGTGTTCTCTTACGGTTCTGGCTTGGCT 1218

Qy   339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
Db   1219 GCCACGGCTCTGTACTCCCTTAAAGTCACACAAGACGCCACACCAGGGAGCGCCCTTGATAAA 1278

RESULT 13
BC031363

LOCUS   BC031363   3278 bp   mRNA   linear   ROD 07-OCT-2003
DEFINITION Mus musculus 3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1, mRNA
            (cDNA clone MGC:36633 IMAGE:5357048), complete cds.
ACCESSION   BC031363
VERSION   BC031363.1 GI:21618632
KEYWORDS   MGC.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 3278)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   MEDLINE 22388257
    PUBMED 12477932
REFERENCE   2 (bases 1 to 3278)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (06-JUN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Contact: Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

FEATURES   Location/Qualifiers
source   1..3278
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                /clone="MGC:36633 IMAGE:5357048"
                /tissue_type="Eye, retina, mouse strain C57B1\6"
                /clone_lib="NIH_MGC_94"
                /lab_host="DH10B"
                /note="Vector: pCMV-SPORT6"
gene   1..3278
                /gene="Hmgcs1"
                /note="synonym: MGC36662"
                /db_xref="LocusID:208715"
                /db_xref="MGI:107592"
CDS   81..1643
                /codon_start=1
                /product="3-hydroxy-3-methylglutaryl-Coenzyme A synthase
                1"
                /protein_id="AAH31363.1"
                /db_xref="GI:21618633"
                /db_xref="LocusID:208715"
                /translation="MPGSLPLNAEEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                GKYTIGLGQARMGFCPTDREDINSLCPTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
                SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                AIYATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTHMQHAVDFYKPDMLSEYPVVDG
                KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFNQKTKAS
                LIVSNQNGNMYTSSVYSSLASVLAQYSPQOLAGKRVGVFSYGSGLAATLYSLKVTQDA
                TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHLANYIPQCSIDSLFE
                GTWYLVRVDEKHRRTYARRPFNDHSLDEGMLVHSNVTATEHIPSPAKKVRPLPATSA
                ESESAVSNGEH"
misc_feature   117..1487
                /note="HMG_CoA_synt; Region:
                Hydroxymethylglutaryl-coenzyme A synthase"
                /db_xref="CDD:pfam01154"

ORIGIN

Alignment Scores:
Pred. No.:   1.66e-190
Score: 2344.00
Percent Similarity: 89.23%
Percent Similarity: 86.92%
Best Local Similarity: 93.35%
Query Match: 10

Length: 3278
Matches: 452
Conservative: 12
Mismatches: 14
Indels: 42
Gaps: 1

US-10-622-516-2 (1-478) x BC031363 (1-3278)

Qy   1   MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
Db   1   ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCTTGCTGGCCAAAAGATGTGGGAATCGTT   60

Qy 21   AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61   GCCCTTGAAATCTACTTTCCTTCTCAATATGTCGATCAAGCTGAGTTGGAAAAATACGAT 120

Qy 41   GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaArgMetGlyPheCysThr   60
         ||||||||||||||||||||||||||||||||||||||| |||||||||||:||||||||
Db 121   ATAACAGCAAGTTTAGGTGACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGGCACCG 180

```
RESULT 14
CHKHMGCOAS
LOCUS       CHKHMGCOAS              1874 bp    mRNA    linear   VRT 06-MAR-1995
DEFINITION  Chicken HMG-CoA synthase mRNA, complete cds.
ACCESSION   M60657
VERSION     M60657.1  GI:211930
KEYWORDS    3-hydroxy-3-methylglutaryl-CoA synthase; HMG-CoA synthase.
SOURCE      Gallus gallus (chicken)
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1874)
AUTHORS     Kattar-Cooley,P.A., Wang,H.H., Mende-Mueller,L.M. and Miziorko,H.M.
TITLE       Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes
            encode the cholesterogenic and ketogenic isozymes
JOURNAL     Arch. Biochem. Biophys. 283 (2), 523-529 (1990)
MEDLINE     91112272
PUBMED      1980405
COMMENT     Original source text: Chicken liver, cDNA to mRNA, clone NC-9.
FEATURES             Location/Qualifiers
     source          1..1874
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9031"
                     /clone="NC-9"
                     /tissue_type="liver"
                     /tissue_lib="Clontech"
     gene            1..1874
                     /gene="HMG-CoA synthase"
     CDS             50..1618
                     /gene="HMG-CoA synthase"
                     /codon_start=1
                     /product="3-hydroxy-3-methylglutaryl-CoA synthase"
                     /protein_id="AAA62737.1"
                     /db_xref="GI:211931"
                     /translation="MPGSLPVNTESCWRKDVGIVALEIYPSQYVDQTELEKYDGVDA
                     GKYTIGLGQSKMGFCSDREDINSLCLTVVQKLMERNSLSYDCIGRLEVGTETIIDKSK
                     SVKTVLMQLFEESGNTDVEGIDTTNACYGGTAALFNAINWIESSMDGRYALVVAGDI
                     AVYATGNARPTGGAGAVAMLVGSNAPLIFERGLRGTHMQAYDFYKPDMVSEYPVVDG
                     KLSIQCYLSALDRCYSVYRNKIHAQWQKEGTDRGPFLLNDFGFMIFHSPYCKLVQKSVA
                     RLLNDFLSDQNAETPANGVVPSGLEAFRDVKLEDTYFDRDVEKAFMKASAELFNQKTKA
                     SLLVSNQNGNMYTPSYVGCLLASLLAQYSPEHLAQGRISEFSYGSGFAATLYSIRVTQD
                     ATPGSALDKITASLSDLKARLDSRKCIAPDVFAENMKIRQETHHLANYIPQCSVEDLF
                     EGTWYLVRVDEKHRRTYARRPVMGDGPLEAGVEVVHHPGIVEHHIFSPAKKVPRIPATT
                     ESEGVTVAISNGVH"
ORIGIN
```

```
US-10-622-516-2 (1-478) x CHKHMGCOAS (1-1874)

  Best Local Similarity:   76.63%
  Query Match:             83.51%
  DB:                      5

  Mismatches:  42
  Indels:      44
  Gaps:        3

Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          |||||||||||||||||||||:::|||||||||||||||||||||:|||||
Db      1 ATGCCCGGGTCTCTTCCTGTGAACACTGAAAGCTGCTGGCCCAAAAGATGTGGGGTATTGTT 109

Qy     21 AlaLeuGluIleTyrProPheProSerGlnTyrValAspGlnThrGluLeuGluLysTyrAsp 40
          |||||||||||||||||||||||||||||||:|||||||||||:|||||||||
Db    110 GCACTGGAAATCTATCCCTCTCAGTATGTCGACCAGACTGAACTGGAGAAGTATGAC 169

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCys 60
          ||||||||||||||||||::|||||||||||||||::::|||||:||||
Db    170 GGTGTGGATGCAGGCAAATACACAATACACGTTAGGCCAGTCAAAAGATGGGGTTTCTGCT 229

Qy     61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
          :::||||||||||||||||||||||||:::|||||||||||||||||||||
Db    230 GACCGAGAGGATATCAATTCCCTCTGTTTGACCGTTGTTCAGAAGCTTATGGAACGGAAC 289

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
          |||||||||||||||||||:||||||||||||||||||||||||
Db    290 AGCCTTTCCTATGACTGCATTGGGAGACTGGAAGTTGGAACGGAGACAATAATTGATAAA 349

Qy    101 SerLysSerValLysThrValLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
          |||||||||||||||||||:|||||||||||:||||||||||||
Db    350 TCAAAATCGGTGAAGACTGTCCTGATGCAGCTATTTGAAGAATCAGGTAATACAGATGTA 409

Qy    121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaIleAsn 140
          ||||||||||||:||||||||||||||||||||||||||||||||||:
Db    410 GAAGGAATTGACACCACAAATGCGTATTGTGGAGGCACACTGGCTCTTTTTAATGCTATT 469

Qy    141 AsnTrpIleGluSerSerSerTrpAsp 149
          ||||||||||||||:||:|||
Db    470 AACTGGATTGAGTCCCAGTTCTTGGACGGCCTATGGACACTTGTGTTGTTGCTGGAGACATT 529

Qy    149 ----------------------------- 149

Db    530 GCTGTGTATGCCACTGGAAATGCCAGGCCAACAGGTGGAGCTGGGGCTGTCGTTGTATGCTA 589

Qy    150 --------------GlyLeuArgGlyThrHisMetGlnHis 158
                        |||||||||||||||||||
Db    590 GTTGGGTCAAATGCTCCTTTAATTTTGAGAGGAGGATTGCGTGGAACCCATATGCAGCAT 649

Qy    159 AlaTyrAspPheTyrLysProAspMetValSerGluTyrProValValAspGlyLysLeu 178
          |||||||||||||||||||||:|:|||||||||||:|||||
Db    650 GCTTATGACTTCTATAAACCAGATATGGTTTCTGAATATCCTGTAGTTGACGGCAAACTA 709

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          |||||||:|||:||||||||||||||||||:|||||||:
Db    710 TCTATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATAGTGTTTATCGCAATAAAATC 769

Qy    199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAsnAspLysPheLeuLeuAsnAspPhe 218
          |||||||||||||:||:|||||||||||||||:|:
Db    770 CATGCCCAGTGGCAGAAAGAGGGACAGGACAGAGGTTTCACCTTGAATGATTTTGGATTC 829

Qy    219 MetIleProPheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
          |||||||||||:||||||||||||||||:::||||||
Db    830 ATGATCTTTCATTCTCCCTACTGTAAACTGGTTCAGAAGTCGGACGACTCTTGCTG 889

Qy    239 AsnAspPheLeuAsnAsnAspGlnAsnArgAsp---LysAsnAsnSerIleTyrSerGlyLeuGlu 257
          |||||||||::|||||||||||:   :|:
Db    890 AATGACTTTCTCAGTGACCAGAATGCAGAAACAGCAAATGGTGTTGTTCCAGTGGGTCTGGAA 949

Qy    258 AlaPheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPhe 277
          |||||:|||||||||||||:|:||||||::|||||||||:
Db    950 GCTTTCGGAGATGTAAAGCTTGAAGACACATATTTTGATAGGGATGTGGAAAAAGCTTTT 1009

Qy    278 MetLysAlaSerAlaGluLeuPheAsnGlnLysThrLysAlaSerLeuLeuValSerAsn 297
          ||||||:||||||||:|||||||||||:||||||||:|||||::||
Db   1010 ATGAAAGCTAGTGCAGAGCTCTTCAATCAGAAAACCAAAGCTTCCTTACTTGTGTCCAAT 1069
```

```
Qy    379 AspValPheAlaGluAsnMetLysIleArgGluAspThrHisHisLeuAlaAsnTyrIle 398
          ||||||||||||||||||||||:|||||||||||||||||||||||||:|||
Db   1341 GATGTCTTTGCTGAAAACATGAAGCTCAGAGAGGACACACATCATTTAGCCAACTATATT 1400

Qy    399 ProGlnGlySerIleAspSerLeuPheGluGlyThrThrTrpTyrLeuValArgValAsnGlu 418
          |||||||||||:||||||||||||||||||||||||||||||:
Db   1401 CCCCAGTGTTCAATAGACTCACTCTTTGAAGGAACGTGGTATCTGGTCAGAGTGGATGAA 1460

Qy    419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
          ||||||||||:|||:::||||:|:|||||||||:|||||||||::
Db   1461 AAACACAGAAGGACATTACGCCCGGGGCCCCTTCACAAATGACCACAGTTTGGATGAAGGA 1520

Qy    439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLeuTyrLysVal 458
          :::|||||||||||||||||||||:||:||||:||||||:::
Db   1521 ATGGGGCTCGTGCATAGTAACATCAGCAACAGAGCATATTCCAAGCCCCTGCTTAAGAAGTG 1580

Qy    459 ProArgLeuProAlaThrAlaLeuGluProGluAlaAlaValIleSerAsnGlyValHis 478
          ||||||||||||||||:|||||||||:::|||||||:::
Db   1581 CCAAGACTCCCCTGCAACCTCGCCGAATCAGCTGTCATCAGTAACGGGAGGAGCAC 1640
```

RESULT 15

```
LOCUS       I65510               1824 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 1 from patent US 5668001.
ACCESSION   I65510
VERSION     I65510.1  GI:2482080
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1824)
  AUTHORS   Miziorko,H.M.
  TITLE     3-hydroxy-3-methyl-glutaryl-coa synthase preparation with improved
            stability
  JOURNAL   Patent: US 5668001-A 1 16-SEP-1997;
FEATURES             Location/Qualifiers
     source          1..1824
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN
```

Alignment Scores:

```
Pred. No.:
Score:                    4.8e-169
Score:                    2089.00
Percent Similarity:       83.33%
Percent Identity:         76.44%
Best Local Similarity:    83.19%
Query Match:
DB:

Length:      1824
Matches:      399
Conservative:  36
Mismatches:    43
Indels:        44
Gaps:           3
```

US-10-622-516-2 (1-478) x I65510 (1-1824)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal  20

Db    2 ATGGCTGGGGTCTTCAGTTACGGAAGCGGACTGGCCACTCTGTACTCTGAAATCCTGTGCCCAAAAGATGTGGGGTATTGTT  61
        |||   |||   |  ||   ||| ::||

Qy   21 AlaLeuGluIleTyrPheProSerGlnThrTyrValAlaAspGlnAlaGluLeuGlyLysTyrAsp  40
Db   62 GCACTGGAAATCTATTTTCCCTCTCAGACTTATGTGGCTGACCAGGCTGAGAAGTGATGAC 121

Qy   41 GlyValAspAlaGlyLysLeuTyrThrArgIleGlyTyrLeuGluGlnAlaLysLysMetGlyPheCysThr  60
Db  122 GGTGTGGATGCAGGCAAATTACACACATTGGGTTACCTTGGAGCCCAGTCAAAAGATGGGCTTCTGCTCT 181

Qy   61 AspArgArgGluAlaIleLeuAsnSerLeuCysMetThrValValAlaGlnAsnLeuGluMetGluArgAsn  80
Db  182 GACCGGAGAGAGGCAATCCTAAATTCCCTCTGTATGACTGTGGTTGCCCAGAACTTAGAGATGGAGAGGAAC 241

Qy   81 AsnLeuSerTyrArgAspCysTyrIleGlyIleArgLeuGluIleGlyThrGluThrIleLeuAspAsnLys  100
Db  242 AGCCTTTCCTATAGGGACTGCTATATTGGGGAGACTGAAGTTGGACGGACACTAATTGATAAA 301

Qy  101 SerLysSerValLysTyrThrAsnLeuMetGlnLeuPheGluGlySerGlyAsnThrAspIle  120
Db  302 TCAAAATCGGTGAAACTGTACCTGATGCAGCTATTTGAAGGAAATCGGTAATACAGATGTA 361

Qy  121 GluGlyIleAspGlyThrArgAsnAlaCysTyrGlyGlyThrAlaAlaLeuAlaPheAsnAlaVal  140
Db  362 GAAGGAAATTGACACAACGCAATGGCTTGCTTCTGCTAGTGGAGCCAGCTGCTTTTTAATGCTATT 421

Qy  141 AsnTrpIleGluSerSerSerTrpAsp  149
Db  422 AACTGGATTGAGTCCAGTTCTTTGGGATGG 481

Qy  150 ----------GlyLeuArgGlyThrHisMetGlnHis  158
Db  482 GCTGTGTATCCCACTGGAAATGCCAGGGCCAACAGTGTGGACTGGTGTGTATGCTA 541

Qy  159 AlaTyrAspPheTyrLysProAspMetLeuSerGlyTyrProLeuValAspGlyLysLysLeu  178
Db  542 GTTGGGTAAATGCTCCTTTAATTTTGAGAGGAGAATTGCGGTGGAACCCACATGCAGCAT 601

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle  198
Db  602 GCTAATGACTTCTATAAACCAGATATGGTTTCTGAATATCCTGTAGTTGATGGCAACTA 661

Qy  199 HisAlaGlnThrTrpGlnLysSerGlyLysAspAspLysPheThrLeuAsnAspPheGlyPhe  218
Db  662 TCTATACAGTGCTACCTCAGCTCAATTAGACCGCTGCTATAGTGTTTATCCGAAATAAATC 721

Qy  219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerGlyAlaArgMetLeuLeu  238
Db  722 CATGGCCCAGTGCAAAAAAGAGGACAGAAGCAATGGGTGTCACCTTGTAAAGATTTTGGATTC 781

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgArgAsp---LysAsnSerIleTyrSerGlyLeuGlu  257
Db  782 ATGATCTTTCATTCTCCCTACTGTAAACTGGTGCAGAAGTCGGTGGCAGACTGTGCTG 841

Qy  258 AlaPheGlyAspValLeuAspThrTyrArgPheAspAspAspAspValGlyLysLysAlaPhe  277
Db  842 AATGACTTTCTCAGTGACCAGAACAGACGAGAAAATGGGTTTTCAGTTGGTTCTGGA 901

Qy  278 MetLysAlaSerGluLeuPheSerGlnLysTyrGlyAlaSerLeuLeuValSerAsn  297
Db  902 GCTTTCAGGGGATGTAGAAGACTTGAAGGATACATATATTTGGAATGCTTTT 961

Qy  298 GlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGln  317
Db  962 ATGAAAGCTAGTGCAGGACTTCTTCAATCAGAAAAACCAAGCCTTCTACTTGTCAGT 1021

Qy  318 TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeu  337
Db 1022 CAGAATGGGAAACATGTACACGTCCTTGCCTTCGTTCTTCTTAGCCCAG 1081

Qy  318 TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeu  337
Db 1130 TACTCTCCGAGACACCTTGCAGGACAAAGAATCGGAGTTCTCAATATGGCTCTGGTTTT 1189

Qy  338 AlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAsp  357
Db 1190 GCTGCTACCGCTGTATTCCATCAGAGTTACACAGATGTTCTGCGCTTGAC 1249

Qy  358 LysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAla  377
Db 1250 AAAATAACTGCTACCCTTTCTGACTTAAAGCAGAGACTTGACTCAGAACGCAATTGCA 1309

Qy  378 ProAspValPheAlaGlyAlaAsnMetLysLeuGluAspThrHisHisLeuValAsnTyr  397
Db 1310 CCTGATGTCTTTGCTGGAGCAAATATGAAATTAGAAGACACACATCATTTGGTCAACTAT 1369

Qy  398 IleGluProGlnGlyLeuPheGluGlyThrThrTrpTyrLeuValArgArgLeu  417
Db 1370 ATTCCACAGGGTTCAGTAGGAGAGATCTCTTGTGAGGGGAACATGG 1429

Qy  418 GlyLysHisArgArgThrTyrAlaAspArgProThrProAsnAspAspThrLeuAspGlu  437
Db 1430 GGAAAACACAGGAGAACATATGCACGACCCCACTCCTAATGGTGATGATACCCTGGACGA 1489

Qy  438 GlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLys  457
Db 1490 GGAGTTGAAGTTGTCCACCCAGCATTGTTCATGACGACCATCCCAAGCCCTGCTAAGAAA 1549

Qy  458 ValProArgLeuProGluAlaThrAla---AlaGluProGlnLeuAlaValIleSerAsnGly  476
Db 1550 GTGCCAAGAATCCCTGCAACAACAGAATCTGAAGGCGTTACTGTTCGCCATTTCAATGGG 1609

Qy  477 GluHis  478
Db 1610 GTGCAT 1615
```

```
Db  1082 TACTCTCCAGAGCACCTTGCAGGAGGACAAAGAATCAGTGAGTTGTTCATATGGCTCTGGTTTT 1141

Qy   338 AlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAsp 357
         |||||||||||:||||:::||:|||||||||||||||||||||||||||||||||||
Db  1142 GCTGCTACGCTGTATTCCATCAGAGTTACACAGGATGCCACTCCTGGTTCTGCGCTTGAC 1201

Qy   358 LysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAla 377
         ||||||||||||:|||||::|||
Db  1202 AAAATAACTGCTAGCCTTTCTGATCTTAAAGCAAGACTTGACTCACCGAAAATGCATTGCA 1261

Qy   378 ProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyr 397
         ||:||:||||:||||||||||||:||:|||:|||||||||||||||
Db  1262 CCTGAGTGTCTTTGCTGAAAACATGAAGATTAGACAGGAGACACATCACTTGGCCAACTAT 1321

Qy   398 IleProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValAlargValAsp 417
         ||||||||||||||||:|::||:
Db  1322 ATTCCACAGTGTTCAGTAGAAGATCTCTTTGAGGGAACATGGTATCTTGTGCGTGTGGAT 1381

Qy   418 GluLysHisSerArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlu 437
         ||||||||||||||||||||||:||:|||
Db  1382 GAAAAACACAGAGAACATATGCACGACGCCCAGTTATGGGTGAATGGACCCCTGGAGGCA 1441

Qy   438 GlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLys 457
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1442 GGAGTTGAAGTTGTCCACCCAGGCATTGTTCATGAGCACATCCCAAGCCCTGCTAAGAAA 1501

Qy   458 ValProArgLeuProAlaThrAla---AlaGluProGluAlaAlaValIleSerAsnGly 476
         |||:|||||||||||||||||      :|:|||
Db  1502 GTGCCAAGAATCCCTGCAACACAGAATCTGAAGGGCGGTTACTGTGTTGCCATTCCAATGGG 1561

Qy   477 GluHis 478
         |||
Db  1562 GTGCAT 1567
```

Search completed: June 24, 2004, 15:10:38
Job time : 4667 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 13:21:38 ; Search time 497 Seconds
                 4085.792 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV..........PRLPATAAEPEAAVISNGEH 478

Searched: 3373863 seqs, 212409041 residues

Scoring table:
                 BLOSUM62
                 Xgapop   10.0 , Xgapext 0.5
                 Ygapop   10.0 , Ygapext 0.5
                 Fgapop   6.0 , Fgapext 7.0
                 Delop   6.0 , Delext   7.0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/Cgn2_1/USPTO_spool_p/US10622516//runat_23062004_162632_639/app_query.fasta_1.647
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10622516 @CGN 1 1 708 @runat_23062004_162632_639 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   N_Geneseq_29Jan04:-
                 1:    geneseqn1980s:*
                 2:    geneseqn1990s:*
                 3:    geneseqn2000s:*
                 4:    geneseqn200las:*
                 5:    geneseqn200lbs:*
                 6:    geneseqn2002s:*
                 7:    geneseqn2002as:*
                 8:    geneseqn2002bs:*
                 9:    geneseqn2003cs:*
                10:    geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2511	100.0	2002	6	ABS54409	Human hyd
2	2511	100.0	2002	7	ABX93299	cDNA enco
3	2480	98.8	1564	7	ACC62337	Human NOV
4	2480	98.8	1564	7	ACC62339	Human NOV
5	2480	98.8	1601	7	ACC62331	Human NOV
6	2480	98.8	1650	7	ACC62336	Human NOV
7	2480	98.8	1650	7	ACC62328	Human NOV
8	2480	98.8	1650	7	ACC62326	Human NOV

ALIGNMENTS

9	2480	98.8	1650	7	ACC62330	Human NOV
10	2480	98.8	1650	7	ACC62338	Human NOV
11	2480	98.8	1650	7	ACC62334	Human NOV
12	2480	98.8	1650	7	ACC62327	Human NOV
13	2480	98.8	1650	7	ACC62332	Human NOV
14	2480	98.8	1650	9	ADE38394	Human pro
15	2480	98.8	3722	10	ADE76933	Human CDN
16	2475	98.6	1593	7	ACC62329	Human src
17	2475	98.6	1608	7	ACC62341	Human src
18	2475	98.6	1608	7	ACC62333	Human NOV
19	2475	98.6	1612	7	ACC62340	Human NOV
20	2475	98.6	1612	7	ACC62335	Human NOV
21	2439	97.1	1685	9	ADD14748	Gene #381
22	2395.5	95.4	2540	5	AAS84743	DNA encod
23	2362	94.1	3275	8	ACF25333	Rat cytos
24	2362	94.1	3275	4	ADB58295	Toxicity-
25	2362	94.1	3275	9	ADB52844	Primary r
26	2089	83.2	1824	2	AAT89089	Avian 3-h
27	2031	65.0	3008	9	AAH34834	Human col
28	1631	65.0	2058	6	ABL69551	Prostate
29	1513	60.3	2058	6	ABN97315	Prostate
30	1513	60.3	2058	6	ACC50182	Breast ca
31	1513	60.3	2058	9	ADB75349	Prostate
32	1513	60.3	2088	4	AAK52002	Human pol
33	1513	60.3	2377	9	ADB36675	Human pro
34	1513	60.2	1994	4	AAH22426	Rat mitoc
35	1511.5	60.2	1994	7	ABT41772	Toxicity-
36	1511.5	60.2	1994	4	ADB57988	Toxicity-
37	1511.5	60.2	1994	9	ADB52471	Primary r
38	1511.5	60.1	2132	9	ADE09760	Novel DNA
39	1510	59.7	2053	3	AAK52986	Human pol
40	1497.5	59.6	2093	3	AAC98123	Human col
41	1337	53.2	1896	4	ABL07457	Drosophil
42	1337	53.2	1939	4	ABL18435	Drosophil
43	1337	53.2	4921	4	ABL07456	Drosophil
44	1337	53.2	6522	4	ABL18434	Drosophil
45	1336	53.2	1752	9	ADE07588	Novel cod

RESULT 1
ABS54409

ID   ABS54409 standard; cDNA; 2002 BP.

AC   ABS54409;

DT   22-NOV-2002 (first entry)

DE   Human hydroxymethylglutaryl-coenzyme A synthase protein cDNA.

KW   Human; gene; ss; chromosome 5; hydroxymethylglutaryl-coenzyme A synthase;
KW   HMG-CoA synthase; cholesterologenesis; therapeutic; diagnostic; genotype;
KW   antibody; synthase; carcinoma.

OS   Homo sapiens.

FH   Key   Location/Qualifiers

FT   5'UTR   1..144
FT   /*tag= a
FT   CDS   145..1581
FT   /*tag= b
FT   /product= "Hydroxymethylglutaryl-coenzyme A synthase"
FT   variation   replace(717,T)
FT   /*tag= c
FT   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR   1582..2002
FT   /*tag= d

PN   US6366692-B1.

PD   20-AUG-2002.

XX

PF 29-MAR-2001; 200IUS-00819993.
XX
PR 29-MAR-2001; 200IUS-00819993.
XX
PA (APPL-) APPLERA CORP.
XX
PI Gong F, Yan C, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-689940/74.
DR P-PSDB; ABG32726.
XX
PT New isolated nucleic acid molecule encoding hydroxymethylglutaryl-CoA synthase, useful as model for the development of human therapeutic targets and for identifying therapeutic proteins.
XX
PS Claim 1; Fig 1; 62pp; English.
XX

CC The invention discloses an isolated nucleic acid molecule encoding a
CC human hydroxymethylglutaryl-coenzyme A (HMG-CoA) synthase which is
CC important in cholesterologenesis. The polynucleotide and polypeptide are
CC useful as models for the development of human therapeutic targets, to aid
CC in the identification of therapeutic proteins and as targets for the
CC development of human therapeutic agents that modulate the activity of the
CC polypeptide in cells and tissues. The polynucleotide is useful for
CC monitoring the effectiveness of modulating compounds on the expression or
CC activity of the enzyme gene in clinical trials and in treatments, in
CC diagnostic assays for qualitative changes in expression of enzyme nucleic
CC acid, to detect mutations in enzyme genes and gene expression products,
CC such as mRNA, for testing an individual for a genotype and as a diagnostic
CC target that can be used to tailor treatment in an individual. The
CC polypeptide is useful to raise antibodies, as a reagent in assays
CC designed to quantitatively determine levels of the protein in biological
CC fluids, as markers for tissues in which the corresponding protein is
CC preferentially expressed (e.g. carcinomas), in biological assays related
CC to members of the synthase subfamily, in drug screening assays, in
CC competition binding assays, in cell-based or cell-free systems, in
CC pharmacogenomic analysis and for treating a disorder characterised by an
CC absence of, inappropriate, or unwanted expression of the polypeptide. The
CC sequence presented is the human HMG-CoA synthase protein cDNA, the gene
CC for which is located on chromosome 5.
XX
SQ Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:

Score:	8.72e-253	Length:	2002
Percent Similarity:	100.00%	Matches:	478
Best Local Similarity:	100.00%	Conservative:	0
Query Match:	100.00%	Mismatches:	0
DB:	6	Indels:	0
		Gaps:	0

US-10-622-516-2 (1-478) x ABS54409 (1-2002)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal  20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAGAGTGTGGGAATTGTT 204

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGluLeuGluLysTyrAsp        40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  205 GCCCTTGAGATCTATTTTCCTTCTCAATATGTGGATCAAGAGCTTGAGAAATATGAT     264

Qy   41 GlyValAspProAlaGlyLysLysTyrThrThrIleGlyLeuGlyGlnAlaLysMetGly  60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  265 GGTGTAGATCCTGGAAAAGTATTACCATTGGCTGGCCCAAGATGGGGTTTCTGCACA     324

Qy   61 AspArgGlyLeuAsnIleAsnSerLeuCysMetThrArgValValGlnAsnLeuMetGlu  80
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  325 GATACAGAGAGATAATTAACTACTCTCTTTGCATGACTCTGGTTCAGAATCTTATGGGA   384

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGlyValGlyThrGluThrIleIleAspLys 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  385 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAA 444

Qy  101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  445 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 504

Qy  121 GluGlyIleAspThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal     140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  505 GAAGGAATCGACACAAACGCTTGCTATGGAGGCACAGCTGCGCTCTTCAATGCTGTT     564

Qy  141 AsnTrpIleGluSerSerSerTrpAspGlyArgTyrGlyLeuValValSerGlyAspTyr 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  565 AACTGGATTGAGTCCAGCTCCTGGGATGGCAGGTACGGCCTTGTGGTCAGCGGTGATTAT 624

Qy  161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  625 GATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAACTCTCCATA 684

Qy  181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysTyrIleHisAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  685 CAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACTGCAAAAAGTACATCCATGCC 744

Qy  201 GlnTrpGlnLysTyrGlyAsnAspPheSerPheThrLeuAlaAspAsnAspPheMetIle 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  745 CAGTGGCAGAAGTACGGGAAATGATTCAGCTTTACCTTGGCAGATAATGATTTCATGATC 804

Qy  221 PheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  805 TTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTGAATGAC 864

Qy  241 PheLeuAsnAspIleGlnAsnArgAspLysGlyAsnSerIleTyrSerGlyLeuAlaPheGly 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  865 TTCCTTAATGACATAGAGAATAGAGATAAAGGTAATTCTATCTATAGTGGCCTGGCATTTGGG 924

Qy  261 AspValLysAspLeuAspThrTyrPheAspAspArgAspValGlyLysAlaPheMetLysAla 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  925 GATGTTAAAGATTTAGAAGACACCTACTTTGAATGATGGAAGGGCATTTATGAAGGCT     984

Qy  281 SerSerGluLeuPheSerGlnLysThrLysSerAlaSerLeuLeuValSerAsnGlnAsnGly 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  985 AGCTCTGAACTCTTCAGTCAGAAAACAAAGTCAGCATCTTTACTTGTATCAAATCAAAATGGA 1044

Qy  301 AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerPro 320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1045 AATATGTACACATCTTCAGTATATGGTTCCCTGGCATCTGTTCTAGCACAGTACTCACCT 1104

Qy  321 GlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr 340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1105 CAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTTCAGGCCTTGCTGCCACT 1164

Qy  341 LeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerProLeuAspLysLeuThr 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1165 CTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTCCGCTCTTGATAAAATAACA 1224

Qy  361 AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspVal 380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1225 GCAAGTTTATGTGATCTTAAATCAAGGCTGGATTCAAGAACTGGTGTGGCACCAGATGTC 1284

Qy  381 PheLeuAlaGluAsnMetLysLeuArgGlyAspThrHisHisSerProSerProAlaLeuGln 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1285 TTCCTGGCTGAAAACATGAAGCTCAGAGGAGACACCCATCATTCGCCATCTCCCGCAG     1344

Qy  401 GlySerIleAspSerLeuGlnPheGluGlyTyrTrpTyrLeuValAlaArgValAspLysHis 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1345 GGTTCAATAGATTCACTCTTTGAAGGATACTGGTACTTAGTTGGTGGGTGGATGAAAGCAC 1404

Qy  421 ArgLysThrTyrAlaLysArgProThrProAsnAsnAspProLeuAspGluGlyValGly 440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1405 AGAAGAACTTACGGTCCGGGTCCACTCCAAAATGATGACCCTTTGGATGAAGGAGTAGGA 1464

Qy  441 LeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysLysValProArg 460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1465 CTTGTGCATTCAAACATAGGAACTCAACATATACCAAGCCCTGCCAAGAAAGTACCAAGA 1524

Qy  461 LeuProAlaThrAlaAlaGluProGluValAlaValIleSerAsnGlyGluHis        478
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1465 CTTGTGCATTCAAACATAGGAACTCAACATATACCAAGCCCTGCCAAGAAAGTACCAAGA 1524
```

```
Db   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   1525 CTCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTTGTCATTAGTPAATGGGGAAACAT 1578

RESULT 2

ID   ABX93299 standard; cDNA; 2002 BP.
XX
AC   ABX93299;
XX
DT   05-JUN-2003 (first entry)
XX
DE   cDNA encoding human HMG-CoA synthase-like enzyme.
XX
KW   Human; cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase;
KW   HMG-CoA synthase; immune response; drug screening assay;
KW   pharmacogenomic analysis; chromosome 5; gene; 88.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   5'UTR          1..144
FT                  /*tag= a
FT   CDS            145..1581
FT                  /*tag= b
FT                  /product= "HMG-CoA synthase-like enzyme"
FT   3'UTR          1582..2002
FT                  /*tag= c
XX
PN   US2002173018-A1.
XX
PD   21-NOV-2002.
XX
PF   12-JUL-2002; 2002US-00193295.
XX
PR   29-MAR-2001; 2001US-00819993.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Gong F, Yan C, Di Francesco V, Beasley EM;
XX
PS   Claim 22; Fig 1A; 65pp; English.
XX
CC   The present invention relates to the isolation of a novel human enzyme
CC   that shows a high degree of similarity to human cytoplasmic 3-hydroxy-3-
CC   methylglutaryl coenzyme A (HMG-CoA) synthase. The invention also
CC   discloses polynucleotide sequences encoding the novel enzyme of the
CC   invention. Both the polypeptide and polynucleotide sequences are useful
CC   as models for the development of human therapeutics, for identifying
CC   therapeutic proteins, as targets for development of human therapeutic
CC   agents, and as query sequences to perform a search against sequence data
CC   bases to identify other family members of related sequences. The
CC   polypeptide is useful to raise antibodies or to elicit another immune
CC   response, as a reagent in assays designed to quantitatively determine
CC   levels of the protein in biological fluids, as markers for tissues in
CC   which the corresponding protein is preferentially expressed, in drug
CC   screening assays, in cell-based or cell-free systems, to identify
CC   compounds that modulate synthase activity of the protein in its natural
CC   state, or an altered form that causes the specific disease or pathology
CC   associated with the synthase, to screen a compound for the ability to
CC   stimulate or inhibit interaction with the synthase protein, and in
CC   pharmacogenomic analysis. The polynucleotide is useful for monitoring the
CC   effectiveness of modulating compounds on the expression or activity of
CC   the human synthase gene in clinical trials or in a treatment regimen, in
CC   diagnostic assays for qualitative changes in a human synthase nucleic
CC   acid that leads to a pathology, for testing an individual for a genotype
CC   that while not necessarily causing a disease, nevertheless affects the
CC   treatment modality, and as antisense constructs to control human synthase
CC   gene expression in cells, tissues and organisms. The present sequence
CC   encodes human HMG-CoA synthase-like enzyme. The gene encoding the enzyme
CC   is located on chromosome 5
XX
SQ   Sequence 2002 BP; 590 A; 380 C; 472 G; 560 T; 0 U; 0 Other;
```

US-10-622-516-2 (1-478) x ABX93299 (1-2002)

Alignment Scores:

```
Pred. No.:
Score:                  8.72e-253    Length:        2002
Percent Similarity:     100.00%      Matches:        478
Percent Identity:       100.00%      Conservative:     0
Best Local Similarity:  100.00%      Mismatches:       0
Query Match:            7            Indels:           0
DB:                                  Gaps:             0
```

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysProLysAspValGlyIleVal 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 204

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  205 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 264

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  265 GGTGTAGAATGCCTGGAAAGTATACCATTGGCTTGTTGAAGGCCCAAGGAAGCCTTCGCA 324

Qy   61 AspArgGluAspIlleAsnSerLeuCysMetThrValGlyPheGluSerTyrGlyGlyAsn 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  325 GATAGAGAAGATATTAACTCTTCTTGCATGACTGTTGGTTCAGAATCTTACGGAGAGAAT 384

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyGlyThrGlyTyrAspProLys 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  385 AACCTTTCCTATGATTGCATTGGGCGGGCCTGGAAGTTGGACAGAGACAACATCACACAA 444

Qy  101 SerLysSerValLysThrArgAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrArgIle 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  445 TCAAAGTCTCTGAAGACTAATTTGATGGAGAGTCGTGAAGAGTCTTGGGAATACAGATATA 504

Qy  121 GluGlyIleIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaIleVal 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  505 GAAGGAATCGACAACAGATAATTTGAGTCAGCCTCGTTGCTGTGTCTTCAAATGCTGTT 564

Qy  141 AsnTrpIleGluSerSerGlnProAspGlyLeuArgGlyLeuGluArgGlyThrHisMetGlnHisAlaTyr 160
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  565 AACTGGATTGAGTCCCAGCCTCTTGGGGATGGGCTTCGTGGGGACACATATGCAACATGCCTAT 624

Qy  161 AspPheTyrLysProAspMetLeuSerGluGlyTyrProIleValAspGlyLysLeuSerIle 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  625 GATTTTTACAAGCCTGATATGGCTAATGCAATCCATAAGTAGGATGGAAAACTCTCCATA 684

Qy  181 GlnCysTyrLeuSerLeuAlaLeuAspArgCysTyrSerValTyrCysGlyLysSerIleHisAla 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  685 CAGTGTGCTACCTCAGTGCATTAGACCGGCTCGTACTCTGTCTCAAAAAGAGATCCAAGCC 744

Qy  201 GlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  745 CAGTGGCAGAAAGAGGGAAAATGAAAAGATTTTACCTTGAATGATTTTGCTTCATGATC 804

Qy  221 PheHisSerProTyrCysLysLeuValGlnLysSerGlnLeuAlaArgMetLeuLeuAsnAsp 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  805 TTTCACTCACCATATGTAAAACTGGTTCAGAAATCTCAGCTCGGATGTTGCTGAATGAC 864

Qy  241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAsnAsp 260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  865 TTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAATGAC 924

Qy  261 AspValLysGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAla 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  925 GATGTTAAAGTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATGAAGGCT 984
```

```
Qy    281  SerSerGluLeuPheSerGlyLysThrLeuLysAlaSerLeuLeuValSerAsnGlnAsnGly  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    985  AGCTCTGAAACTCTTCAGTCAGGAAAACAAAGGCATCATTTACTTGTATCAAATCAAAATGGA  1044

Qy    301  AsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnThrYrSerPro  320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1045  AATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGACTTCACCT  1104

Qy    321  GlnGlnLeuAlaGlyLysArgIleGlyValPheSerGlyLysValLeuAlaAlaThr  340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1105  CAGCAATTAGCAGGGAAGAGAAATTGGAGTGTTTTCTTATGGTTTGGCTGCCCACT  1164

Qy    341  LeuTyrSerLeuLysValThrGlnAspAlaThrProGlyGlySerAlaLeuLysLysIleThr  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1165  CTGTACTCTCTTAAAGTCACACAGGATGCTACACCGGGGGTCTCTGCCTTGATAAAAATAACA  1224

Qy    361  AlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaLaProAspVal  380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1225  GCAAGTTTATGTGATCTTAAATCAAGGCTTGAACCTAGGACTGGTGTGGCACCAGATGTC  1284

Qy    381  PheAlaGluGluAsnMetLysLeuArgGluArgGlnAspThrHisHisLeuValAsnTyrIleProGln  400
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1285  TTCGCTGAAGAAAACATGAAGCTCAGAGAGCGACACCCATCATTTGGTCAACTATATTCCCCAG  1344

Qy    401  GlySerIleAspSerSerLeuPheGluGlyTyrTrpTyrLeuValAlaArgValAlaAspGluGlyLysHis  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1345  GGTTCAATAGATTCACGTCTTTGAGGGAACGTGGTACTTAGTTAGGGGTGGCGATGAAAAGCAC  1404

Qy    421  ArgArgThrTyrAlaArgArgProThrProAsnAspAspAspAspThrLeuAspGlyLyValValGly  440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1405  AGAAGAACTTACGCTCGGCGTCCCACTCCCAAATGATGACGATGACACTTTGGATGGAAGGTGTTAGGA  1464

Qy    441  LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysGlyLysValProArg  460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1465  CTTGTGCATTCAAACAATAGCAAACATGGCATTAATTCCAAGCCCTGCCAAGAAGGTACCAAGA  1524

Qy    461  LeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerArgAsnGlyGluHis  478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1525  CTCCCGCCCGCCACAGCAGCAGAACCTGAGGCAGCTGCTGTCATTAGTAATAATGGGGGAACAT  1578
```

Alignment Scores:

Pred. No.:	
Score:	1.08e-249
Percent Similarity:	2480.00
Best Local Similarity:	91.92%
Query Match:	91.92%
DB:	98.77%

Length:	1564
Matches:	478
Conservative:	0
Mismatches:	42
Indels:	1
Gaps:	1

US-10-622-516-2 (1-478) x ACC62337 (1-1564)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaGlnCysTrpProLysAspValGlyIleVal  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTCAGTGCTGGCCAAAAGATGTGGGAATTGTT  61

Qy   21 AlaLeuGluIleTyrPheProSerGlnThrTyrValAlaAspGlnAlaGluLeuGluLysTyrAsp  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 GCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGCACACAGAGTTGGAAAAATATGAT  121

Qy   41 GlyValAspAlaGlyLysLysTyrThrIleGlyLeuGlnAlaLysMetGlyPheCysThr  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 GGTGTAGATGCTGGAAAGTATACCATTGGGCTGCAGGCCAAGATGGGCTTCTGCACA  181

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTTGTTCAGAATCTTATGGAGAGAAAT  241

Qy   81 AsnLeuSerTyrAspCysSerIleGlyTyrArgLeuValGlyThrGlyThrIleIleLeuAspGlu 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  242 AACCTTTCCTATGATTGCAGTATTGGTTACCGCCTTGTGGGAACAGGAACAATCATCCTTGACGAA  301

Qy  101 SerLysSerValLysThrAsnLeuPheGluSerGlyTyrAsnThrAspIle 120
          ||||||||||||||||||||||||||||||||||||||||||
Db  302 TCAAAGTCTGTGAAGACTAATTTGTTTGAGAGTGGTTATAACACAGATATA  361

Qy  121 GluGlyIleAspThrThrAsnAlaCysTyrGlyValGlyThrAlaAlaAlaValPheAlaAsnAlaVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  362 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCCGCTGCTGTCTTCAATGCTGTT  421

Qy  141 AsnTrpIleGluSerSerSerTrpAsp  149
          
Db  422 AACTGGATTGAGTCCAGCTCTTGGGATGGGGTATGCCCTGGTAGTTGCAGGAGATATT  481

Qy  149 ------------------  149
          
Db  482 GCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGCAGTAGCTCTGCTA  541

Qy  150 ------------------GlyLeuArgGlyThrHisMetGlnHis  158
                             |||||||||||||||||||
Db  542 ATTGGGGCCAAATGCTCCTTTAATTTTGAACGAGCGGGCCTTCGTGTGGGACACATATGCAACAT  601

Qy  159 AlaThrTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysVal  178
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  602 GCCACTTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC  661

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle  198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  662 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTTTACTGCAAAAAGATC  721

Qy  199 HisAlaGlnIleTrpGlnLysGlyAsnAspLysAsnPheThrLeuAsnAspPheGlyPhe  218
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  722 CATGCCCAGTGGCAGAAAGGAAATGATAAAAACTTTACCTTGAATGATTTTGGCTTC  781

Qy  219 MetIlePheHisSerProTyrCysGlySerValGlnLysSerLeuAlaArgMetLeuLeu  238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  782 ATGATCTTTCACTCACCATATTGTGAACGTGTTCAGAAAATCTCAGCTCGGAATGCTTG  841

Qy  239 AsnAspPheLeuAsnAsnGlnAsnAspTyrValLeuIleTyrSerGlyTyrLeuGluAla  258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  842 AATGACTTTCCTTAATGACCAGAATAGAGATAAAAATAGTTAACTATATAGTGCCTGAAGCC  901

Qy  259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysValAlaPheMet  278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  902 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG  961

Qy  279 LysAlaSerGluLeuPheSerGlnLysThrLysGlyAlaSerLeuLeuValSerAsnGln  298
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  962 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTGTATCAAATCA  1021

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr  318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1022 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC  1081

Qy  319 SerProGlnIleGlnLeuAlaGlyLeuGlyValPheSerTyrIleGlyValPheSerGly  338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1082 TCACCTCAGCAATTAGCAGGGACTTGGTGTTTTCTTATGGTTATGGTTGGCT  1141

Qy  339 AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1142 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTTCTGCCTTGATAAA  1201

Qy  359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyAlaPro  378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1202 ATAACAGCAAGTTTATGTGATCTTAAAAATCAAGGCTTAGTTCAAGAAACTGGTGCACCA  1261

Qy  379 AspValPheAlaGlyLeuAsnMetLysSerLeuGluAspThrHisHisLeuValAsnTyrIle  398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1262 GATGTCTTCGCTGGAACAATGAAGTCTCAGAGACCAGGACACCATCATTGGTCAACTATATT  1321

Qy  399 ProGlnSerIleLeuAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu  418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1322 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTTGATGAA  1381

Qy  419 LysHisArgArgThrTyrArgLeuArgProThrProAsnAspAspProThrLeuAspGlyGly  438
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1382 AAGCACAGAAGAACTTACCGTCTGCGGCCCACTCCAAATGACGATCCAACACTTTGATGAAGGA  1441

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleLeuProSerProAlaLysLysVal  458
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1442 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATATTCCAAGCCCCTGCCAAGAAAGTA  1501

Qy  459 ProArgGluLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis  478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1502 CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT  1561
```

RESULT 4

ACC62339

ID	ACC62339 standard; cDNA; 1564 BP.
XX	
AC	ACC62339;
XX	
DT	23-JUN-2003 (first entry)
XX	
DE	Human NOV4in encoding cDNA SEQ ID NO:207.
XX	
KW	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW	haemostatic; immunomodulator; antiinflammatory; antiasthmatic; anti-HIV;
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW	immune disorder; haematopoietic disorder; dyslipidaemia;
KW	metabolic syndrome X; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200302301-A2.
XX	
PD	20-MAR-2003.
XX	
PF	09-SEP-2002; 2002WO-US028538.

```
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322817P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0323519P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323336P.
PR   20-SEP-2001; 2001US-0323361P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2001US-0355999P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
PA
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP,  Anderson DW,   Berghs C,    Boldog FL;
PI   Burgess CE, Casman SU;  Catterton E,   Chant JS,    Chaudhuri A;
PI   Crabtree J, Dipippo VA; Ediger SR,     Eisen AJ,    Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L,       Gorman L,    Guo X, Gusev VV, Ji W;
PI   Kekuda R,  Khramtsov NV,  Leach MD,     Lepley DM,  Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE,        Ort T,      Padigaru M, Patturajan M;
PI   Pena CEA,  Rieger DK,    Rothenberg ME, Shenoy SG,  Shimkets RA;
PI   Spadena SK, Spytek KA,   Taupier RJ,    Twomlow N,  Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54270.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 246-247; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   associated with the NOVX protein in humans and for treating a syndrome
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC   disorders, neoplasm, lymphoma, uterus cancer, diabetes, metabolic
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1564 BP; 464 A; 304 C; 362 G; 434 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                   1.08e-249
Score:                       2480.00
Percent Similarity:          91.92%       Length:              1564
Best Local Similarity:       91.92%       Matches:             478
Query Match:                 98.77%       Conservative:        0
DB:                                       Mismatches:          0
                             7            Indels:              42
                                          Gaps:                1

US-10-622-516-2 (1-478) x ACC62339 (1-1564)
```

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaAlaCysTrpProLysAspValGlyIleVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGCTGGCCCAAAAGATGTGGGAATTGTT 61

Qy     21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuLeuGlyLysTyrAsp 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAATTATGAT 121

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyLeuLeuGlyGlnAlaLysMetGlyPheCysThr 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GGTGTAGATGCTGGAAAGTATACCATTGGCCTTGGGCAGGCCAAGATGGGCTTCTGCACA 181

Qy     61 AspArgGluAspIleAsnAsnLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 241

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleGluAspLys 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAA 301

Qy    101 SerLysSerValLysThrAsnLeuMetCInLeuPheGluGluLysGlyTyrAsnThrAspIle 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTTGGATACAGATA 361

Qy    121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaAlaValPheAsnAlaVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGCTGTGTTCAATGCTGTT 421

Qy    141 AsnTrpIleGluSerSerSerTrpAsp------------------------------------ 149
          |||||||||||||||||||||||||||
Db    422 AACTGGATTGAGTCCAGCTCTTGGGAT------------------------------------ 481

Qy    149 --------------------------------------------------------------- 149
          
Db    482 GCTGTATATGCCACAGGAAATGCTAGACCTACACAGGTGGGAGTTGGAGCAGTAGCTCTGCTA 541

Qy    150 --------GlyLeuArgGlyThrHisMetGlnHis 158
                  ||||||||||||||||||||||||||
Db    542 AATTGGGCCAAATGCTCCTTTAATTTTGACGAGGGCTTCGTGGGACACACATATGCAACAT 601

Qy    159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyGlyLeu 178
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAGGAACTC 661

Qy    179 SerIleGlyCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 TCCATACAGGTGCTACCTCAGTGCATTAGACCGCTGCTACTCGTCTACTGCAAAAAGATC 721

Qy    199 HisAlaGlnTrpGlnGlyGlyAsnAspPheThrLeuAsnAspPheGlyPhe 218
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CATGCCCAGTGGCAGAAAGGAGGAAATGATTTTACCTTGAATGATTTTGGCTTC 781

Qy    219 MetIlePheHisSerProTyrCysLysLeuValGlnLeuSerLeuLeuAlaArgMetLeuLeu 238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    782 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCCTAGCTCGGATGTTGCTG 841

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    842 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTTATATGTGGGCCTGGAAGCC 901

Qy    259 PheGlyAspValLysGluAspThrTyrPheAspArgAspValGlyLysAlaPheMet 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    902 TTTGGGGATGTTAAAATTAGAAGACACCTACTTTGATAGAGATGTGGGAAGGCCATTATG 961
```

```
Qy    279 LysAlaSerGluLeuPheSerGlyLysThrLysAlaSerLeuLeuValSerAsnGln 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    962 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAAACAAAAGGCATCTTTACTTGTATCAAATCAA 1021

Qy    299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuValAlaSerValAlaGlnTyr 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1022 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCAGTCGTTCTAGACAGTAC 1081

Qy    319 SerProGlnGlnLeuAlaGlyIleGlyValPheSerTyrGlySerGlyLeuVal 338
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1082 TCACCTCAGCAATTAGCAGGGAGAAGAATTGGAGTGTTTTCTTATGGTTTGGCT 1141

Qy    339 AlaThrLeuTyrSerLeuLysValThrGlnAsnAlaThrProGlySerAlaLeuAspLys 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1142 GCCACTCTGTACTCTCTTAAAGTCACACAGAAGCTACACCGGGGTCTGCTCTTGTATAAA 1201

Qy    359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1202 ATAACAGCAAGTTTATGTGATCTTAAAATCAAGACTTGATAGCAGGACTGGTGTGGCACCA 1261

Qy    379 AspValPheAlaGluAsnMetLeuArgGluArgHisHisLeuLeuValAsnTyrIle 398
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1262 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGAGACACCACATCATTTGGTCAACTATATT 1321

Qy    399 ProGlnGlySerIleAspSerLeuPheGluGluLeuTyrThrTrpTyrLeuValLysArgValAspGlu 418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1322 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTGTAGGGGTGGATGAA 1381

Qy    419 LysHisArgGlyThrTyrAlaArgGlyProThrProAsnAspAspThrGluGlu 438
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1382 AAGCACAGAGGAACTTACGCTCGGGGTCCACTCCCAAATGATGACACTTTGGAGGA 1441

Qy    439 ValGlyLeuValHisSerAsnIleAlaThrGlnHisIleProSerProAlaLysGlyLysVal 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1442 GTAGGACTTGTGCATTCAAACATAGCAACTCAGCATATTCCAAGCCCTGCCAAGGGAAAGTA 1501

Qy    459 ProArgLeuProAlaThrAlaAlaGluProGlyAlaAlaValIleSerAsnGlyGluHis 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1502 CCAAGACTTCCCTCCCACAGCAGCAGAACCTGAAGCAGCAGTTGTCAATTAGTAATGGGGAACAT 1561
```

```
XX
ID   ACC62331 standard; cDNA; 1601 BP.
XX
AC   ACC62331;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV4lf encoding cDNA SEQ ID NO:191.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; 88.
XX
OS   Homo sapiens.
XX
PN   WO2003023001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
```

```
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   20-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0324999P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324999P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00226417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli BA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54262.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 242-243; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating in humans and for treating
CC   associated with a NOVX protein in a human disease. NOVX nucleic acids,
CC   and antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1601 BP; 473 A; 321 C; 369 G; 438 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:   1.12e-249
Score: 2480.00
Percent Similarity: 91.92%
Best Local Similarity: 91.92%
Query Match: 98.77%
DB:

Length: 1601
Matches: 478
Conservative: 0
Mismatches: 0
Indels: 42
Gaps: 1

US-10-622-516-2 (1-478) x ACC62331 (1-1601)
DB:

Qy	1 MetProGlySerLeuProLeuAsnAlaGluAsnAlaGlnValAlaCysTrpProLysAspValGlyTyrIleVal 20
Db	1 ATGCCCGGATCACTTCCTTTGAAATGCAGAAGCAGAAGCTGGAAATTGTT 72
Qy	21 AlaLeuGlyIleLeuTyrPheProSerGlnThrTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
Db	13 GCCCTTGGACATCTATTTTCCTTCCAATAGTGTACGTCAATAACGTGGAAAAATATGAT 132
Qy	41 GlyValAspAlaProAlaGlyLysTyrThrIleGlyLeuGlyLnAlaAlaLysMetGlyPheCysThr 60
Db	73 GGGTGTAGATGCTGGAAAGTATACAATTGGCTTGGGCAAGCTGCAGCGTTCTGCACA 192
Qy	61 AspAlaArgGlyAspIleAsnSerLeuCysSerMetThrValValGlnAsnLeuMetGluArgAsn 80
Db	193 GATAGAGAAGGATAATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 252
Qy	81 AsnLeuSerTyrAspCysIleGlyTyrArgLeuValGlyThrGluThrIleIleIleAspLys 100
Db	253 AACCTTCCTATGATTGCATTGGGCGCTGGAAGTTGGAACAGAGAACAAATCATCGACAA 312
Qy	101 SerLysSerValLysThrAsnLeuMetGluLeuPheGluGluSerGlyAsnThrAspIle 120
Db	313 TCAAAGTCTGTGAAGACTAAATTGAATGCAGTTGTTTGAAGAGTCAAATACAGATATA 372
Qy	121 GluGlyIleLeuAspThrThrArgAsnAlaCysTyrArgGlyGlyThrArgAlaAlaValPheAsnAlaVal 140
Db	373 GAAGGAATCGACACAACTAAATGCATGCTTATGGAGGCACAGCTCCTGTCTTCAAATGCTGTT 432
Qy	141 AsnTrpIleGluSerSerSerTrpAsp 149
Db	433 AACTGGATTGAGTCCAGCTCTTGGGATGGGTATGCCCTGGTAGTTGCAGGAGATATT 492
Qy	149 --------- 149
Db	493 GCTCGTATATGCCACAGAGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 552
Qy	150 ---------GlyLeuArgGlyThrHisMetGlnHis 158
Db	553 ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGAGGCTTCGTCGTGGACACATAAGCAACAT 612
Qy	159 AlaTyrArgAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyIleLysLeu 178
Db	613 GCCTATAGAGATTTTTACAAGCCCGATATGCTATCTGAATATCCTATAGTAGATGGGATAAAACTC 672
Qy	179 SerIleGlnCysTyrLeuSerArgLeuAspArgCysTyrArgSerValTyrCysLysLysSerIle 198
Db	673 TCCATACAGTGCTACCTGAGTCGATTAGACCGCTGCTACCGTAGTGTACTGCAAAAAGATC 732
Qy	199 HisAlaGlnIrrpGlyIleGlnGlyAsnAspAspPheThrLeuAsnAspPheGlyPhe 218
Db	733 CATGCCCAGTGGCGAAAGCAGAGGGAAATGAATAAAGATTTTACCTTGAATGATTTTGGCTTC 792
Qy	219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
Db	793 ATGATACTTTCACTCCACCATATTGTAAGCTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 852
Qy	239 AsnAspPheLeuAsnAsnArgAspGlyAsnSerIleTyrSerGlyLeuGluAla 258
Db	853 AATGACTTCCTTAATGACCAGAATAGAGATGAAAAAAGTAGTATCTATAGTGGGCCTGGAAGCC 912
Qy	259 PheGlyAspValLysLeuGluLeuAspThrThrTyrPheAspArgAspValGluLysAlaPheMet 278
Db	913 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGAGATGTGGAGAAGGCATTTATG 972

RESULT 6

ACC62336
ID ACC62336 standard; cDNA; 1650 BP.
XX
AC ACC62336;
XX
DT 23-JUN-2003 (first entry)
XX
DE Human NOV4lk encoding cDNA SEQ ID NO:201.
XX
KW Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW haemostatic; antiinflammatory; anti-HIV; immunomodulator;
KW neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW congenital heart defect; aortic stenosis; valve disease; transplantation;
KW tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW fertility; haemophilia; hypercoagulation; graft versus host disease;
KW idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW immune disorder; haematopoietic disorder; dyslipidaemia;
KW metabolic syndrome X; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200023001-A2.
XX
PD 20-MAR-2003.
XX
PF 09-SEP-2002; 2002WO-US028538.

```
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322826P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323635P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381489P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PA   Novel human proteins and nucleic acid encoding the proteins, useful for
PA   diagnosis, treatment and prevention of disorders involving the human
PA   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 245; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABRS4167 to ABRS4276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent NOV sequences and probes for human NOV
CC   sequences, which are used in examples from the present invention.
CC   ABRS4277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:                1.17e-249
Score:                    2480.00
Percent Similarity:       91.92%
Best Local Similarity:    91.92%
Query Match:              98.77%
DB:
```

```
US-10-622-516-2 (1-478) x ACC62336 (1-1650)

                 Length:        1650
                 Matches:       478
                 Conservative:  0
                 Mismatches:    0
                 Indels:        42
                 Gaps:          1
```

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysGlyGlyIleVal 20
        |||
Db    1 ATGGCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGGATGTTGGGATTGTT 81

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlyAsnThrAspIle 40
        |||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATGGAAATACAGATATA 141

Qy   41 GlyValAspAlaGlyLysLysTyrThrIleLeuGlyLysGlnAlaLysMetGlyPheCysThr 60
        ||||||||||||||||||||||||||||||||||||||||||
Db  142 GGTGTAGATGCTGGGAAGTATACCATTGGGAGGCCAAGATGGGCTTCTGCACA 201

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuAsnLeuMetGluArgAsn 80
        |||||||||||||||||||||||||||||||||||||||||
Db  202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTAATGCTGTT 261

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrHisIleIleAspLys 100
        |||||||||||||||||||||||||||||||||
Db  262 AACCTTTCCTATGATTGCATTGGGCGGCTTGAAGTTGGAACAGAGACAATCATCGACAAA 321

Qy  101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGlnSerGlyAsnThrAspIle 120
        |||||||||||||||||||||||||||||||||
Db  322 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 381

Qy  121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
        ||||||||||||||||||||||||||||||
Db  382 GAAGGAAATCGACACAAACTAATGCAAGCTATGGGAGGCCACACACTGCCGTGTCTTCAATGCTGTT 441

Qy  141 AsnTrpIleGluSerSerSerTrpAsp 149
        ||||||||||||||||
Db  442 AACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTGGCAGGGAGATATT 501

Qy  149 --------------------------------------------- 149
Db  502 GGCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGGGACCAGTAGGCTCTGCTA 561

Qy  150 --------------------GlyLeuArgGlyThrHisMetGlnHis 158
                            ||||
Db  562 ATTTGGGGCCAAATGCTCCCTTTAATTTTGAACGAGGGCTTCGTGGGCACACATATGCAACAT 621

Qy  159 AlaTyrAspPheTyrLysProAspMetLeuSerGlyTyrProIleValAspGlyLysLeu 178
        |||||||||||||||||||||||||||||||||
Db  622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGGATATCCTATAGTAGATGGGAAACTC 681

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
        |||||||||||||||||||||||||||||||||
Db  682 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGTATTCTGTTCTACTGCAAAAAGATC 741

Qy  199 HisAlaGlnTrpGlnLysGlyAsnAspLysAspGluLysPheGlyPhe 218
        ||||||||||||||||
Db  742 CATGCCCAGTGGCAGAAGGGAAATGATAAAGATGAAAAATTTGGCTTC 801

Qy  219 MetIlePheHisSerProTyrCysLysLysValGlnLysSerLeuAlaArgMetLeuLeu 238
        |||||||||||||||||||||||||||||||||
Db  802 ATGATCTTCACTCACCATATGTAAACTGGTTCAGAAAATCTCAGCTCCGGATAGTTGCTG 861

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
        |||||||||||||||||||||||||||||||||
Db  862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC 921

Qy  259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysAlaPheMet 278
        |||||||||||||||||||||||||||||||||
Db  922 TTTGGGGATGTTAAATTAGAAGACACCTACTTCGATAGAGATGTGGGAAAGGCATTATG 981
```

```
Qy  279  LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln  298
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  982  AAGGCTAGCTCCTGAACTTCTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAA  1041

Qy  299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr  318
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCAATCTGTTCTAGCACAGTAC  1101

Qy  319  SerProGlnGlnLeuAlaGlyIleGlyValPheSerTyrGlySerGlyLeuAla  338
              | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1102 TCACCTCAGCAATTAGCAGGAAGAGAATTGGAGTGGTTTTCTTATGGTTCTGGTTTGGCT  1161

Qy  339  AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1162 GCCACTCCTGTACTTCTTTAAAGTCACACAAGATGGACTGCTCTCTTGATAAA  1221

Qy  359  IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro  378
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1222 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGGGTGGCCACCA  1281

Qy  379  AspValPheAlaLeuAsnMetLysSerAsnGluArgGluAspThrHisHisLeuValAsnTyrIle  398
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1282 GATGTCTTCGCTGAAAAACATGAAGCTTCAGAGAGGGACACCCATCATTTGGTCAACTATATT  1341

Qy  399  ProGlnGlySerIleAspAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu  418
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1342 CCCCAGGGTTCAATAGATGATAGTCTTTTGAAGGGAACGTGGTACTTAGTGGGATGGA  1401

Qy  419  LysHisArgGlnThrTyrAlaArgArgProThrProAsnAspAspAspThrLeuAspGluGly  438
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1402 AAGCACAGACAGACAACTTACGCTCGGGGTCCCAATCCCAAATGATGATGACACTTTGGATGAAGGA  1461

Qy  439  ValGlyLeuValHisSerAsnIleAlaThrGluHisIleGluProSerProAlaLysLysGlyVal  458
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1462 GTAGGACTTGTGCATTCAAACATAGCCACTGAGCATATTGAACCCAGCCCTGCCAAGAAAGTA  1521

Qy  459  ProArgGlyLeuProAlaThrAlaAlaGluProGlyAlaAlaValIleIleSerAsnGlyGluHis  478
              | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1522 CCAAGGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT  1581
```

```
RESULT 7
ACC62328
ID   ACC62328 standard; cDNA; 1650 BP.
XX
AC   ACC62328;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV4c encoding cDNA SEQ ID NO:185.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder X; gene; ss.
OS   Homo sapiens.
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Guoev VV, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2001US-0359599P.
PR   05-MAR-2002; 2001US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396402P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
WPI; 2003-313241/30.
P-PSDB; ABR54259.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 241; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62245 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunomodulator, neuroprotective,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, anorexia, cancer-associated cachexia,
CC   infectious disease, Crohn's disease, multiple sclerosis, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	1.17e-249	Length:	1650
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Best Local Similarity:	91.92%	Mismatches:	0
Query Match:	98.77%	Indels:	42
		Gaps:	1

US-10-622-516-2 (1-478) x ACC62328 (1-1650)

```
Qy      1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
Db      1  ATGCCTGGATCACTTCCTTTGAATGCACAAGCTTGCTGGCCAAAAGATGTTGGGATTGTT   81

Qy     21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlyLysTyrAsp   40
Db     82  GCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGACCAAGCAGAGTTGGGAAAATATGAT  141

Qy     41  GlyValAspAlaGlyLysTyrThrIleLeuGlyLeuGlyAsnAlaLysMetGlyPheThr   60
Db    142  GGTGTAGATGCTGGGAAGTATACCATCCTGGGACTGGGAAATGCAAGATGGGCTTCACA   201

Qy     61  AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
Db    202  GATAGAGAAGATATAAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT  261

Qy     81  AsnLeuSerTyrArgAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAsp  100
Db    262  AACCTTTCCTATAGATTGCATTGGGCGGCTTGGAAGTTGGAACAGAGACAATCATCGACAAA  321

Qy    101  SerLysSerValGlyThrAsnLeuMetGlnLeuPheGluLeuSerGlyAsnThrAspIle  120
Db    322  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAACTGAGTGGAAATACAGATATA  381

Qy    121  GluGlyIleAspThrThrAsnAlaCysTyrGlyIleThrAlaAlaValPheAsnAlaVal  140
Db    382  GAAGGAAATCGACACAACAATAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT  441

Qy    141  AsnTrpIleGluSerSerSerTrpAsp                                   149
Db    442  AACTGGATAGAGTCCAGCTCTTGGGATGGA                                501

Qy    149  ------------------------------                                149
Db    502  GCTGTATATGCCACAGGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTGCTA   561

Qy    150  --------------GlyLeuArgGlyThrHisMetGlnHis                     158
Db    562  ATTGGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT  621

Qy    159  AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLeu     178
Db    622  GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAACTC  681

Qy    179  SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle  198
Db    682  TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC  741

Qy    199  HisAlaGlnIleTrpGlnLysGluGlyAsnAspPheThrLeuAsnAspPheGlyPhe     218
Db    742  CATGCCCAGTGGCAGAAAGAGGGAAATGATACCCTTAACTTGAATGATTTTGGCTTC     801

Qy    219  MetIlePheHisSerProTyrCysGlyLeuValGlnLysSerGluLeuAlaArgMetLeuLeu  238
Db    802  ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTGAATTGGCACGGATGTTGCTG  861

Qy    239  AsnAspPheLeuAsnAspLysAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
Db    862  AATGACTTCCTTAATGACAAGAATAGAGATAAGAACTCTATCTATAGTGGCCTGGAAGCC  921

Qy    259  PheGlyAspValLysLeuGluAspThrTyrPheArgArgAspValGluLysAlaPheMet  278
Db    922  TTTGGGGATGTTAAATTAGAAGACACCTACTTCAGAAGAGATGTGGAGAAGGCATTTATG  981
```

RESULT 8

ACC62326

ID	ACC62326 standard; cDNA; 1650 BP.
XX	
AC	ACC62326;
XX	
DT	23-JUN-2003 (first entry)
XX	
DE	Human NOV4la encoding cDNA SEQ ID NO:181.
XX	
KW	Human; NOVX; antiatherosclerotic; hypertensive; cardiant; dermatological;
KW	anorectic; antiinfective; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW	haemostatic; immunomodulator; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;
KW	idiopathic thrombocytopenic purpura; infectious disease; anorexia;
KW	Crohn's disease; multiple sclerosis; Alzheimer's disease; Parkinson's disease;
KW	cancer-associated cachexia; Alzheimer's disease; haematopoietic disorder; dyslipidaemia;
KW	immune disorder; metabolic syndrome X; gene; ss.
OS	Homo sapiens.
XX	
PN	WO200302300l-A2.
XX	
PD	20-MAR-2003.
XX	
PF	09-SEP-2002; 2002WO-US028538.

XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318404P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0411144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393132P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-0236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Bergha C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spadera SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
WPI; 2003-313241/30.
DR   P-PSDB; ABR54457.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 240; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62245 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC   disorders, neoplasm, lymphoma, uterus cancer, diabetes, metabolic
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62236 to ACC62245 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.17e-249   Length: 1650
Score: 2490.00   Matches: 478
Percent Similarity: 91.92%   Conservative: 0
Best Local Similarity: 91.92%   Mismatches: 0
Query Match: 98.77%   Indels: 42
DB: Gaps: 1

US-10-622-516-2 (1-478) x ACC62326 (1-1650)

```
Qy    1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22  ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTTGGGATTGTT   81

Qy   21  AlaLeuGluIleTyrPheProSerGlnThrTyrValAspGluLeuGluLysGlyTyrAsp   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82  GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATGAACTGGAGAAAGGATATGAT    141

Qy   41  GlyValAspAlaGlyLeuTyrThrIleGlyTyrLeuGlyLeuGlnAlaLysMetGlyCysThr  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142  GGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGGCTTCAGGCCAAGATGGGCTTCTGCACA  201

Qy   61  AspArgGluAspIleAsnSerLeuCysMetThrValIleValGlnAsnLeuMetGluArgAsn  80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202  GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGATTGTTCAGAATCTTATGGAGAGAAAT  261

Qy   81  AsnLeuSerTyrArgAspCysIleGlyArgLeuGlnValGlyThrGlnThrIleIleAspLys  100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262  AACCTTTCCTATAGATTGCAT...                                        321

Qy  101  SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGlnSerGlyAsnThrAspIle   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322  TCAAAGTCTGTGAAGACTAAT...                                        381

Qy  121  GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyTyrThrAlaAlaValPheAsnAlaVal  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382  GAAGGAAATCGACACAATAATGCA...                                     441

Qy  141  AsnTrpIleGluSerSerSerTrpAsp                                    149
         ||||||||||||||||||||||||||
Db  442  AACTGGATTGAGTCCAGCTCTTGGGATGCACGGTATGCCCCTGGTAGTTGCAGGAGATATT  501

Qy  149  --------------------GlyLeuArgGlyThrHisMetGlnHis                158
                             ||||||||||||||||||||||||||
Db  502  GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA   561

Qy  150  ----GlyLeuArgGlyThrHisMetGlnHis                               
Db  562  ARTGGGCCAAATGCTCCTTTAATTTGAACGAGGGCGCTTCGTGGGACACACATATGCCAGAT  621

Qy  159  AlaThrTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu  178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622  GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAAACTC    681

Qy  179  SerIleGlnCysTyrLeuSerAlaLeuAspAspArgCysTyrSerValTyrCysLysGlyIle  198
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682  TCCATACAGTGCTACCTCAGTGCATTAGACCCGCTATTCTGCTACTCGAAAAGATC       741

Qy  199  HisSerLeuGlnIleTrpGlnLysGlyLysAspSerPheThrLeuAsnAspPheGlyPhe    218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742  CATGCCCAGTGGCAGAAAGGAAAGGATAGATTTACCTTGAATGATTTTGGCTTC         801

Qy  219  MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuValLeuArgMetLeuLeu  238
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802  ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAAATCTCTAGCCTCGGATGTTGCTG  861

Qy  239  AsnAspPheLeuAsnAspArgAspLysAsnSerIleTyrSerGlyLeuGluLeuGluLysAla  258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862  AATGACTTTCTTAATGACCGAGATAAGAATAGTATCTATATAGTGGGCCTGGAAGCC       921

Qy  259  PheGlyAspValLysLeuLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet  278
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922  TTTGGGGATGTTAAATTAGAGACACCTACTTTGATAGAGATGTGGAGAGGCACATTTATG    981
```

```
RESULT 9

Qy    279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    982 AAGGCTAGCTCGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTGTATCAAATCAA 1041

Qy    299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1101

Qy    319 SerProGlnLeuGlnLeuAlaGlyIleGlyValPheSerTyrGlySerGlyLeuLeuAla 338
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1102 TCACCTCAGCAATTAGCAGGAGAAGGAAGTGGAGTGTTTTCTTATGGTTCTGGTTTGGCT 1161

Qy    339 AlaThrLeuTyrSerLeuLysValThrArgGlnAspAlaThrProGlySerAlaLeuAsp 358
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1162 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTTCTGCTCTTGATAAA 1221

Qy    359 IleThrAlaSerLeuCysAspLeuHisLysSerArgLeuAspSerArgThrGlyValAla 378
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 ATAACAGCAAGTTTATGTGATCTTCATAAATCAAGGCTTGATTCAAGAACTGGTGTGGCA 1281

Qy    379 AspValPheAlaGluLeuAsnMetLysSerLeuGlyAspThrHisHisLeuValAsnTyrIle 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 GATGTTTTCGCTGAACTGAATATGAAGTCAGAGGACACCCATCATTTGGTCAACTATATT 1341

Qy    399 ProGlnGlySerIleAspSerLeuPheGluGluValThrTrpTyrLeuValArgValAspGlu 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1342 CCCCAGGGTTCAATAGATTCACTCTTTGAAGAAGTGTTACTTAGTTGAGGTGGATGAA 1401

Qy    419 LysHisArgGlyThrTyrAlaArgProThrProAsnAspAspArgLeuTrpSerGlyGly 438
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1402 AAGCACAGAAGGAACTTACCGCTCCGACTCCAAATGATGACAGACTTTGGAGTGGAAGGA 1461

Qy    439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1462 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521

Qy    459 ProArgLeuProAlaThrAlaAlaGluProGluLeuAlaValIleSerSerAsnGlyGluHis 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1522 CCAAGACTCCCTGCCACAGCAGAGAACCTGAACGAGTCTTGAGTAGTGGGAACAT 1581
```

ID   ACC62330 standard; cDNA; 1650 BP.
XX
AC   ACC62330;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV41e encoding cDNA SEQ ID NO:189.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; 88.
XX
OS   Homo sapiens.
XX
PN   WO200303001-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323536P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0399599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381483P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0364532P.
PR   13-AUG-2002; 2002US-0403512P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI   Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54261.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 242; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, scleroderma, congenital heart defects,
CC   valve disease, tuberous sclerosis, obesity, aortic stenosis,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, anorexia, Crohn's disease, multiple
CC   sclerosis, infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:
Pred. No.:
Score: 1.17e-249
Percent Similarity: 2480.00
Best Local Similarity: 91.92%
Query Match: 91.92%
DB: 98.77%

	Length:	1650
	Matches:	478
	Conservative:	0
	Mismatches:	0
	Indels:	42
	Gaps:	1

US-10-622-516-2 (1-478) x ACC62330 (1-1650)

```
Qy    1  MetProGlySerLeuProLeuAsnAlaGluAlaAlaCysTrpProLysAspValGlyTyrLeuVal  20
Db   22  ATGCCCTGGAGATCACTTCCTTTGAAATGCAGAAGAGAAGATGTTGGGATTGTT  81
```

(sequence alignment continues)

```
Qy   21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp  40
Db   82  GCCCTTGAGATCTATTTTCCTTCCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT  141
```

```
Qy   41  GlyValAspAlaLeuGlyLysTyrThrIleGlyTyrLeuGlyGlnAlaLysMetGlyPheCysThr  60
Db  142  GGTGTAGATGCTGGGAAGTATACACTTGGTTATCTGGGTCAAGCAAAGATGGGCTTCTGCACA  201
```

```
Qy   61  AspArgGlnLeuAspIleAsnAsnLeuCysMetThrValValGlnAlaAsnLeuMetGluArgAsn  80
Db  202  GATAGACAAGAATATTAACTTCTCTTGCATGACAGTCGTGTTCAAGAATCTTATGGAGAAAT  261
```

```
Qy   81  AsnLeuSerTyrAspCysIleGlyTyrArgLeuGluValGlyGlyThrGluThrIleIleIleAspLys  100
Db  262  AACCTTTCCTATGATTGCATTGGGCTACCGGCTGGAAGTAGGAGGGAACAGAGACAATCATCACGACAAA  321
```

```
Qy  101  SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle  120
Db  322  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA  381
```

```
Qy  121  GluGlyIleAspThrThrAsnAlaCysTyrArgGlyGlyThrAlaAlaValPheAsnAlaVal  140
Db  382  GAAGGAAATCGACACAAACTAAATGCATGGGATAGGAGGCACAGCTGCTGTTTCAATGCTGTT  441
```

```
Qy  141  AsnTrpIleGluSerSerSerIleArgAsp------------------  149
Db  442  AACTGGAATAGAGTCCAGCTCTTGGGATGGACCGGTATGCCCTGGTAGTTGCAGGAGATATT  501
```

```
Qy  149  ------------------GlyLeuArgGlyThrHisMetGlnHis  158
Db  502  GCTCGTATATGCCACAGGAGAAATGCTACAGCATTCTGGAGTGGAGCAGTAGCTCTGCTA  561
```

```
Qy  159  AlaTyrAspPheTyrLysProAspMetLeuSerGlyTyrProIleValAlaGlyLysLeu  178
Db  562  ATTGGCGCAAATGCTCCTTTAATTTTGAACGAAGGACTTCGTGGACACAGTAGCAACAT  621
```

```
Qy  179  SerIleGlnCysTyrLeuSerArgLeuAlaLeuAspArgCysTyrSerSerValTyrCysLysLysLysIle  198
Db  622  GCCTATGATTTTTACAAGCCCTGAATGCTATCTGAAATACCTATAGTAGATGGGAAACTC  681
```

```
Qy  199  HisAlaGlnIleTrpGlnLysGluGlyAspGlyAlaAsnAspSerGlyGlyPheGlyPhe  218
Db  682  TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTACTGTACAAAAGATC  741
```

```
Qy  219  MetIlePheHisSerProTyrCysGlyLysLeuValGlnLysSerLeuLeuArgMetLeuLeu  238
Db  742  CATGCCCAGTGCCAGGAAAGATAAGATTTTACCTTGGAATGAATGGGCTTC  801
```

```
Qy  239  AsnAspPheLeuAsnAspGlnAsnArgAspSerGlyGlyLeuGluAla  258
Db  802  ATGATTCTTTCACTCACCATATTGTAAACTGGTTCAGTCTCGGAGTTGCTG  861
```

```
Qy  259  PheGlyAspValLysGluGlyAsnThrTyrPheAspArgAspValGlnLysAlaPheMet  278
Db  862  AATGACTTCCTTAATGACCAGAATAGAGATAGTGGTGGCCTGGAAGCC  921
```

```
Qy  279  LysAlaSerSerGluLeuPheSerGlnGlyLysThrLysAlaSerLeuLeuValSerAsnGln  298
Db  922  TTTGGCGGATGTTAAAATTAGAGAGACACCTACTTTGATAGAGATGTGGAGAGGCATTTATG  981
```

```
Qy  299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuValAlaSerValLeuAlaGlnTyr  318
Db  982  AAGGCTAGCTCCTGAACTCTTCAGTCAGGGAAGAATTGGAGTGTTTTCTTATAGGGTTCCGT  1041
```

```
Qy  319  SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla  338
Db 1042  AATGGAAAAATATGTACACATCTTCAGTATATATGGTGTTCCCTTGGCATTCGTTGCATTGCTGTCTAGCACAGTAC  1101
```

```
Qy  339  AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
Db 1102  TCACCTCAGCAAATTAGCACGGGAAGAGAAATTGGAGTGTTTTCTTATAGGGTTCCGT  1161
```

```
Qy  359  IleThrAlaSerLeuCysAspSerLeuValSerAsnSerArgLeuAsnSerArgThrGlyValAlaPro  378
Db 1162  GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCTCTTGATAAA  1221
```

```
Qy  379  AspValPheAlaGluAsnMetLysSerLeuValArgGluAspThrHisHisGlnValValAsnTyrIle  398
Db 1222  ATAACAGCAAGTTTAATCGATCTTAAATCAACGGCTTGATTCAAGAAACTGGTGTGCCACCA  1281
```

```
Qy  399  ProGlnGlySerIleAspSerGlyLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu  418
Db 1282  GATGTCTTCGCTGAAAAACATGAAAGCCTCAGAGAGGACACCCATCATTTGGTCAACTATATT  1341
```

```
Qy  419  LysHisGluArgThrTyrAlaArgArgProThrProAsnAspAspAspThrLeuAspGluGly  438
Db 1342  CCCCAGGGTTCAATAGATTCACTCCTTTGAAGGAACCTGGTACTTAGTTAGGGTGGATGAA  1401
```

```
Qy  439  ValGlyLeuValHisSerAsnIleAlaThrGluIleProSerProAlaLysLysVal  458
Db 1402  AAGCACGAGAGAACTTACGGCTCGGCGTCCCACTCCAAATGACACTTTGGATGAGGGA  1461
```

```
Qy  459  ProArgLeuProAlaThrAlaAlaLeuGluProGluAlaAlaValIleIleSerAsnGlyGluHis  478
Db 1462  GTAGGGACTTGTGCATTCAAACATAGCAACTGAGCATNNNNCCAAGCCCTGCCAAGAAAGTA  1521
```

```
Db 1522  CCAAGACTCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTGTCATTAGTTAATGGGGAACAT  1581
```

```
XX
PR   07-SEP-2001; 200US-0318120P.
PR   07-SEP-2001; 200US-0318184P.
PR   10-SEP-2001; 200US-0318430P.
PR   17-SEP-2001; 200US-0324990P.
PR   17-SEP-2001; 200US-0322265P.
PR   17-SEP-2001; 200US-0322781P.
PR   17-SEP-2001; 200US-0322816P.
PR   17-SEP-2001; 200US-0322817P.
PR   19-SEP-2001; 200US-0323519P.
PR   20-SEP-2001; 200US-0323631P.
PR   20-SEP-2001; 200US-0323636P.
PR   25-SEP-2001; 200US-0324969P.
PR   25-SEP-2001; 200US-0325091P.
PR   26-SEP-2001; 200US-0324990P.
PR   14-DEC-2001; 200US-0341144P.
PR   26-FEB-2002; 200US-0359599P.
PR   05-MAR-2002; 200US-0361663P.
PR   03-MAY-2002; 200US-0377908P.
PR   17-MAY-2002; 200US-0381493P.
PR   29-MAY-2002; 200US-0383863P.
PR   02-JUL-2002; 200US-0393332P.
PR   17-JUL-2002; 200US-0396412P.
PR   13-AUG-2002; 200US-0403517P.
PR   06-SEP-2002; 200US-0236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI   Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI   Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI   Gangolli EA, Gerlach VL, Giot L, Gorman L, Guso X, Gusev VY, Ji W;
PI   Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI   Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Pattuzajan M;
PI   Pena CEA, Rieger JM, Rothenberg ME, Shenoy SG, Shimkets RA;
PI   Spadena SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI   Zerhusen BD, Zhong M;
XX
DR   WPI; 2003-113241/30.
DR   P-PSDB; ABR54269.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human NOVX
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 246; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. AC062236 to AC062345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antiispaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acid, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   AC062346 to AC062465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:                     1.17e-249      Length:          1650
Score:                         2480.00        Matches:          478
Percent Similarity:            91.92%         Conservative:       0
Best Local Similarity:         91.92%         Mismatches:        42
Query Match:                   98.77%         Indels:             1
DB:                            7              Gaps:               1
```

US-10-622-516-2 (1-478) x ACC62338 (1-1650)

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysGlyGlyVal  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGGTGGTGTT  81

Qy     21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlyTyrAsp  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGAGCAGGCTGAACTTGGATACGAT  141

Qy     41 GlyLeuAspAlaGlyLysThrTyrThrIleGlyLeuGlnAlaLysMetGlyPheCysThr  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 GGTGTTGGGATCCTGGGAAGTGTACCAATTGGCTTGGGCTTCTGCACA  201

Qy     61 AspArgGluAspIleAsnSerLeuCysSerMetThrValValGlnAsnLeuMetGluArgAsn  80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 GATAGAGAAGATATTAACTCTCTTTGCAGACTGTGGGTTCAGAATCTTATGGAGAGAAAT  261

Qy     81 AsnLeuSerTyrAspCysIleGluGlyArgLeuGluValGlyThrGluThrIleIleIleAlaAspLys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 AACCTTTCCTATGATTGTATTGAAGGGCGTGGAAGTGGAACAGAGACAATCATCGACAA  321

Qy    101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyLysAsnThrAspIle  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 TCAAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAAAAATACAGATATA  381

Qy    121 GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyGlyThrArgAlaAlaValPheAsnAlaVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 GAAGGAATCGACACACAACTAATGCATGCTATGGGAGGGCACAGCTGCTGCTGTCGTTCTTCAATGCTGTT  441

Qy    141 AsnTrpIleLeuGluSerSerSerTrpAsp-----------------  149
          |||||||||||||||||||||||||||||
Db    442 AACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCGTGAGTTGCAGGAGTATT  501

Qy    149 ----------------------------------------  149
Db    502 GCTGTATATCCACAGGAAATGCTAGACCTACAGGTGGGAGTTGGAGACTCTGCTA  561

Qy    150 ---------GlyLeuArgGlyThrHisMetGlnHis  158
                   |||||||||||||||||||||||||
Db    562 ATTGGGCCAAATGCTCCTTTAAATTTTGGAACGAGGGCTTCGTGGGACACATATGCACAT  621

Qy    159 AlaTyrAspPheTyrLysProAspMetLeuSerGlyTyrProIleValAspGlyLysLeu  178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAACTC  681

Qy    179 SerIleGlnTyrLeuSerArgAlaLeuAspAspArgCysTyrSerValTyrCysLysLysIle  198
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 TCCATACAGTGCTACCTCAGTGCATTAGACGCTGCTATTCTGTCTACTCCAAAAAGATC  741

Qy    199 HisAlaGlnTrpGlnLysGlyValAsnAspPheThrLeuAsnAspPheGlyPhe  218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 CATGCCCAGTGGCAGAAAGGAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC  801

Qy    219 MetIlePheHisSerProTyrCysSerLeuValGlnLysSerLeuAlaArgMetLeuLeu  238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 ATGATCTTTCACTCACCATATTGTAGCCTGGTTCAGAAATCTCTAGCCCGGATGTTGCTG  861

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC  921

Qy    259 PheGlyAspValLysGluGluAspThrTyrPheAspArgAspValGlyLysGlyAlaPheMet  278
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGGAAAGGGCATTATG  981
```

```
Qy     279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
Db     982 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAA 1041

Qy     299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
Db    1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1101

Qy     319 SerProGlnGlnLeuAlaGlyLysValPheSerTyrGlySerGlyLeuAla 338
Db    1102 TCACCTCAGCAATTAGCAGGGAAAGTAGAATTCTCTTATGGTTCTGGGTTTGGCT 1161

Qy     339 AlaThrLeuTyrSerLeuGlyValThrGlnAspAlaThrProGlySerAlaLeuAspPro 358
Db    1162 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCCGGGTCTGCTCTTGATAAA 1221

Qy     359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
Db    1222 ATAACAGCAAGTTTAGTGACTTAAATCAAGGCTTGATTCAAGAACTGGGTGTGGCACCA 1281

Qy     379 AspValPheAlaGlnAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
Db    1282 GATGTCTTCGCTCAGAAAACATGAAAGCTCAGAGAGGACACCCATCATTGGGTCAACTATATT 1341

Qy     399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
Db    1342 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGGGTGGATGAA 1401

Qy     419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspAspGluGly 438
Db    1402 AAGCACAGAAGAACTTACGCTCGGCGTCCACTCCAAATGATGACACTTGGATGAAGGA 1461

Qy     439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleLeuProSerProAlaLysLysVal 458
Db    1462 GTAGGACTTGTGCATTCAAACATAGCCAACTGCAGCATAATTCCAAGCCCTGCCAAGAAAGTA 1521

Qy     459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
Db    1522 CCAAGACTCCCTCCACAGCCAGCTGCTGCTGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT 1581
```

```
RESULT 11
ACC62334
ID   ACC62334 standard; cDNA; 1650 BP.
XX
AC   ACC62334;
XX
DT   23-JUN-2003 (first entry)
XX
DE   Human NOV4li encoding cDNA SEQ ID NO:197.
XX
KW   Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW   anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW   haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW   neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW   gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW   congenital heart defect; aortic stenosis; valve disease; transplantation;
KW   tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW   prostate cancer; neoplasm; lymphoma; uterus cancer;
KW   fertility; haemophilia; hypercoagulation; graft versus host disease;
KW   idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW   Crohn's disease; multiple sclerosis; infectious disease; cancer;
KW   cancer-associated cachexia; Alzheimer's disease, Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia;
KW   metabolic syndrome X; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200302301-A2.
XX
PD   20-MAR-2003.
XX
PF   09-SEP-2002; 2002WO-US028538.
XX
```

```
XX
PR   07-SEP-2001; 2001US-0318120P.
PR   07-SEP-2001; 2001US-0318184P.
PR   10-SEP-2001; 2001US-0318430P.
PR   17-SEP-2001; 2001US-0322636P.
PR   17-SEP-2001; 2001US-0322781P.
PR   17-SEP-2001; 2001US-0322816P.
PR   17-SEP-2001; 2001US-0322817P.
PR   19-SEP-2001; 2001US-0323519P.
PR   20-SEP-2001; 2001US-0323631P.
PR   20-SEP-2001; 2001US-0323636P.
PR   25-SEP-2001; 2001US-0324969P.
PR   25-SEP-2001; 2001US-0325091P.
PR   26-SEP-2001; 2001US-0324990P.
PR   14-DEC-2001; 2001US-0341144P.
PR   26-FEB-2002; 2002US-0359599P.
PR   05-MAR-2002; 2002US-0361663P.
PR   03-MAY-2002; 2002US-0377908P.
PR   17-MAY-2002; 2002US-0381433P.
PR   29-MAY-2002; 2002US-0383863P.
PR   02-JUL-2002; 2002US-0393332P.
PR   17-JUL-2002; 2002US-0396412P.
PR   13-AUG-2002; 2002US-0403517P.
PR   06-SEP-2002; 2002US-00236417.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Agee ML,  Alsobrook JP,  Anderson DW,  Berghs C,  Boldog FL;
PI   Burgess CE,  Casman SJ,  Catterton E,  Chant JS,  Chaudhuri A;
PI   Crabtree J,  Dippo VA,  Edinger SR,  Eisen AJ,  Ellerman K;
PI   Gangolli EA,  Gerlach VL,  Glot L,  Gorman L,  Gusev VY,  Ji W;
PI   Kekuda R,  Khramtsov NV,  Leach MD,  Lepley DM,  Guo X,  Liu X;
PI   Malyankar UM,  Miller CE,  Ooi CE,  Ort T,  Padigaru M,  Patturajan M;
PI   Pena CEA,  Rieger DK,  Rothenberg ME,  Shenoy SG,  Shimkets RA;
PI   Spaderna SK,  Spytek KA,  Taupier RJ,  Twomlow N,  Vernet CAM,  Voss EZ;
PI   Zerhusen BD,  Zhong M;
XX
DR   WPI; 2003-313241/30.
DR   P-PSDB; ABR54265.
XX
PT   Novel human proteins and nucleic acid encoding the proteins, useful for
PT   diagnosis, treatment and prevention of disorders involving the human
PT   protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS   Claim 20; Page 244; 460pp; English.
XX
CC   The present invention describes isolated human NOVX proteins, where X is
CC   1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC   ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC   hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC   antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC   antiasthmatic, immunomodulator, neuroprotective, nootropic,
CC   antiparkinsonian and antilipaemic activities, and can be used in gene
CC   therapy. NOVX proteins are useful for treating or preventing a pathology
CC   associated with a NOVX protein in humans and for treating a syndrome
CC   associated with the human disease. NOVX nucleic acids, proteins and
CC   antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC   atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC   valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC   congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC   disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC   hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC   disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC   infectious disease, anorexia, cancer-associated cachexia, cancer,
CC   Alzheimer's disease, Parkinson's disease, immune disorders,
CC   haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC   ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC   sequences, which are used in examples from the present invention.
CC   ABR54277 represents a human trypsinogen protein given in comparison with
CC   the human NOV35b protein in the exemplification of the present invention
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:

Pred. No.:	1.17e-249
Score:	2480.00
Percent Similarity:	91.92%
Best Local Similarity:	91.92%
Query Match:	98.77%

Length:	1650		
Matches:	478		
Conservative:	0		
Mismatches:	0		
Indels:	42		
Gaps:	1		

US-10-622-516-2 (1-478) x ACC62334 (1-1650)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 ATGCCTGGATCACTTCCTTTGAATGCAAAGCTTGCTGGCCAAAAGATGTTGGGATTGTT 81

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluGluGluLysTyrAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GCCCTTGAGATCTATTTTCCTTCCAAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 141

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyGlyValAlaLysMetGlyPheCysThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 GGTGTAGATGCTGGGAAGTATACCATTGGCTGGGTATACCAAGATGGGCTTCTGCACA 201

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 GATAGAGAAGATATAAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 261

Qy   81 AsnLeuSerTyrAspCysIleGlyTyrArgLeuGlyGlyThrGlyThrIleIleIleAsp 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 AACCTTTCCTATGATTGCATTGGATATCGTCTGGGCGGTGGAACAGGAACAATCATCGAC 321

Qy  101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluGlyAsnThrAspIle 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTTGGGAAATACAGATATA 381

Qy  121 GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382 GAAGGAATCGACAACAACTAATGATGCTTATGGAGGCACAGCTGCTGTTCTTAATGCTGTT 441

Qy  141 AsnTrpIleGluSerSerSerTrpAsp 149
          |||||||||||||||||||||||||||
Db  442 AACTGGATTGAGTCCAGCTCTTGGGAT 501

Qy  149 ----------------------------------------- 149
          
Db  502 GCTGTATATGCCACAGGAAATGCTAGACGTACAGGTGGAGTTGGAGCACTAGCTGCTA 561

Qy  150 ----------GlyLeuArgGlyThrHisMetGlnHis 158
                    |||||||||||||||||||||||||||
Db  562 ATGGGGCAAATGCTCCTTAATTTGAACGAGGGCTTCGTGGGACACACATATGCAACAT 621

Qy  159 AlaTyrAspPheTyrLysProAspMetLeuSerGlnTyrProIleVal AspGlyLeu 178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622 GCCTATGATTTTTACAAGCCTGATATGCTATCTCAATATCCTATAGTAGATGGAAACTC 681

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC 741

Qy  199 HisAlaGlnThrTrpGlnLysGlyLysAspAlaProLysAsnPheThrLeuAsnAspPheGlyPhe 218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742 CATGCCCAGTGGCAGAAAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 801

Qy  219 MetIlePheHisSerProTyrCysLysGlyLysValGlnLysSerLeuAlaArgMetLeuLeu 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802 ATGATCTTTCACTCACCATATTGTAAAGGTAAAGTGCAAAAATCTCTAGCTCGGATGTTGCTG 861

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAspValGlyLeuGlyLysAlaPheMet 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862 AATGACTTCCTTAATGACCAGAATAGAGATAAAGATGTTGGACTCGGAAAGGCATTTATG 921

Qy  259 PheGlyAspValLysLeuGluGlyLeuLysSerGlyLeuGluAlaPheMet 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922 TTTGGGGATGTTAAATTAGAGGGACTCAAGTCTGGCCTGGAAGCATTTATG 981

Qy  279 LysAlaSerGluGluPheSerGlnTyrGlyThrLysAspSerLeuLeuValSerAsnGln 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982 AAGGCTAGCTCTGAACTCTTCAGTCAGTATGGAACAAAGGACAGTCTTACTGTATCAAACCAA 1041

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaServalLeuAlaGlnTyr 318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCGTTCTTCAGCACAGTAC 1101

Qy  319 SerProGlnGlnLeuAlaGlyLysLysArgIleGlyLeuValPheSerTyrGlySerGly 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102 TCACCTCAGCAAATTAGCAGGAAGAAAATGGAGTGGTTTTCTTATGGTGTATCAAATATTG 1161

Qy  339 AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162 GCCACTCTGTACTCTCTTGTAACACAAGATGCTACACCGGGGTCTGCTCTTGATAAA 1221

Qy  359 IleThrAlaSerLeuCysAspAspLeuAlaSerArgLeuAspSerArgThrGlyValAlaPro 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222 ATAACAGCAAGTTTATGCGATCTTGCAAGTCGTTAAAATCAAGACTGGTGGCACCA 1281

Qy  379 AspValPheAlaGluAsnMetLysLeuArgIleGlyAspThrHisHisLeuValAsnTyrIle 398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282 GATGTCTTCGCTGAAAACATGAAGCTCAGAAGAGCACCCACATCATTGGTCAACTATATT 1341

Qy  399 ProGlnGlySerIleAspSerSerLeuPheGluGlyThrTrpTyrLeuValArgValGlyGlu 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1342 CCCCAGGGTTCAATAGATTCACTCTTTGAGGGAACGTGGTACTTAGTGAGGGTGGGTGAA 1401

Qy  419 LysHisArgArgThrTyrArgLeuArgArgProThrProAsnAspAspAspThrLeuAspAspGlyGly 438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1402 AAGCACAGAAGAACCTACAGGCTCCGGAGACCAACTCCAAATGAATGACGACACTTTGGATGAAGGA 1461

Qy  439 ValGlyLeuValHisSerAsnIleAlaIleThrGluHisSerIleProSerProAlaLysGlyLysVal 458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1462 GTAGGACTTGTGCATTCAAATATAGCAACTGAGCATATTCCAAGCCCTGCCAAAGGAAAGTA 1521

Qy  459 ProArgLeuProAlaThrAlaAlaGluProGluAlaIleAlaValIleIleSerArgAsnGlyGluHis 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1522 CCAAGACTCCCTGCCACAGCAGCGGAACCTGAAGCAGCTATTGCAGTCATTATAAGTAAGGGAAACAT 1581
```

RESULT 12
ACC62327

ID	ACC62327	standard; cDNA; 1650 BP.
AC	ACC62327;	
DT	23-JUN-2003 (first entry)	
DE	Human NOV41b encoding cDNA SEQ ID NO:183.	
KW	Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;	
KW	anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;	
KW	haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;	
KW	neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;	
KW	gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;	
KW	congenital heart defect; aortic stenosis; valve disease; transplantation;	
KW	tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;	
KW	prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;	
KW	fertility; haemophilia; hypercoagulation; graft versus host disease;	
KW	idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;	
KW	Crohn's disease; multiple sclerosis; infectious disease; cancer;	
KW	cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;	
KW	immune disorder; haematopoietic disorder; dyslipidaemia;	
KW	metabolic syndrome X; gene; ss.	
OS	Homo sapiens.	
XX		
PN	WO200023001-A2.	
XX		
PD	20-MAR-2003.	
XX		
PF	09-SEP-2002; 2002WO-US028538.	

XX
PR 07-SEP-2001; 2001US-0318120P.
PR 07-SEP-2001; 2001US-0318184P.
PR 10-SEP-2001; 2001US-0318430P.
PR 17-SEP-2001; 2001US-0326256P.
PR 17-SEP-2001; 2001US-0322781P.
PR 17-SEP-2001; 2001US-0322816P.
PR 17-SEP-2001; 2001US-0322817P.
PR 19-SEP-2001; 2001US-0323519P.
PR 20-SEP-2001; 2001US-0323631P.
PR 20-SEP-2001; 2001US-0323626P.
PR 25-SEP-2001; 2001US-0324969P.
PR 25-SEP-2001; 2001US-0325091P.
PR 26-SEP-2001; 2001US-0324990P.
PR 14-DEC-2001; 2001US-0411144P.
PR 26-FEB-2002; 2002US-0359599P.
PR 05-MAR-2002; 2002US-0361663P.
PR 03-MAY-2002; 2002US-0377908P.
PR 17-MAY-2002; 2002US-0381483P.
PR 29-MAY-2002; 2002US-0383863P.
PR 17-JUL-2002; 2002US-0396412P.
PR 02-JUL-2002; 2002US-0393332P.
PR 13-AUG-2002; 2002US-0403517P.
PR 06-SEP-2002; 2002US-00236417.
XX

PA (CURA-) CURAGEN CORP.

PI Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI Burgess CE, Casman SJ, Catterton E, Chant JS, Chaudhuri A;
PI Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Gusev VY, Ji W;
PI Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI Zerhusen BD, Zhong M;
XX

DR WPI; 2003-313241/30.
DR P-PSDB; ABR54258.
XX

PT Novel human proteins and nucleic acid encoding the proteins, useful for
PT diagnosis, treatment and prevention of disorders involving the human
PT protein or nucleic acid e.g. cardiac and neurological disorders.
XX

PS Claim 20; Page 240-241; 460pp; English.
XX

CC The present invention describes isolated human NOVX proteins, where X is
CC 1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC hypotensive, dermatological, anorectic, immunosuppressive, cytostatic,
CC antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC antiasthmatic, metabolic, immunomodulator, neuroprotective, nootropic,
CC antiparkinsonian and antilipaemic activities, and can be used in gene
CC therapy. NOVX proteins are useful for treating or preventing a pathology
CC associated with a NOVX protein in humans and for treating a syndrome
CC associated with the human disease. NOVX nucleic acids, proteins and
CC antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC congenital adrenal hyperplasia, prostate cancer, fertility, haemophilia,
CC hypercoagulation, neoplasm, lymphoma, uterus cancer, diabetes, metabolic
CC disorders, idiopathic thrombocytopenic purpura, graft versus host
CC disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC infectious disease, anorexia, cancer-associated cachexia, cancer,
CC Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC sequences, which are used in examples from the present invention.
CC ABR54277 represents a human trypsinogen protein given in comparison with
CC the human NOV35b protein in the exemplification of the present invention
XX

SQ Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;

Alignment Scores:
Pred. No. :   1.17e-249
Score: 2480.00
Percent Similarity: 91.92%
Best Local Similarity: 91.92%
Query Match: 98.77%
DB:

US-10-622-516-2 (1-478) x ACC62327 (1-1650)

Length: 1650
Matches: 478
Conservative: 0
Mismatches: 0
Indels: 42
Gaps: 1

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
Db   22 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTTGGGATTGTT 81

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlyLysTyrAsp 40
Db   82 GCCCTTGAGATCTATTTTCCTTCCCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 141

Qy   41 GlyValAspIleGlyLysTyrThrIleGlyLysGlyThrAlaAlaLysMetGlyPheCysThr 60
Db 142 GGTGTAGATATCGGGAAGTATACCATTGGCTTGGGCCCAGGCCAAGATGGGCTTCTGCACA 201

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValValGlnAsnLeuMetGluArgAsn 80
Db 202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 261

Qy   81 AsnLeuSerTyrAspCysSerLeuGlyArgGlyLeuGluValGlyThrGluThrIleHisMetGlnHis 100
Db 262 AACCTTTCCTATGATTGCAGTCTGGGACGCGGTCTGGAAGTTGGAACAGAGACAATCATCATGCAGCAT 321

Qy 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
Db 322 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 381

Qy 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal 140
Db 382 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCCACAGCTGCTGTTCTTCAATGCTGTT 441

Qy 141 AsnTrpIleGluSerSerTrpAsp----------GlyLeuArgGlyThrHisMetGlnHis 149
Db 442 AACTGGATTGAGTCCAGCTCTTGGGGATGGACGGGTATGCCCTGGTTGCAGGAGATATT 501

Qy 149 --------------------GlyLeuArgGlyThrHisMetGlnHis 149
Db 502 GCTGTATATGCCAGGAAATGCTAGACCTACAGGTGGAGCAGCAGTAGCTCTGCTA 561

Qy 150 ----------GlyLeuArgGlyThrHisMetGlnHis 158
Db 562 ATTGGGCCAAATGCTCCTTTAATTTTGAACGAGGCCGGTTCGTGCGGACACACATATGCACAT 621

Qy 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAlaAspGlyLysLeu 178
Db 622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGCAGATGGAAACTC 681

Qy 179 SerLeuGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerSerValTyrCysLysLysIle 198
Db 682 TCCATACAGTGCTACCTCAGTGCAATAGACCGCTGCTATTCTGTCTATTGCAAAAAGATC 741

Qy 199 HisLeuGlnTrpGlnLeuGlyAspAspSerPheThrLeuAsnAspPheGlyPhe 218
Db 742 CATGCCCAGTGGCAGCTTGGAGATGACAGTTTTACCTTGAATGATTTTGGCTTC 801

Qy 219 MetIlePheHisSerProTyrCysLysLysValGlnLysSerLeuValGlnArgMetLeuLeu 238
Db 802 ATGATCTTTCACTCACCATATTGTAAACAGGTTCAGAAAATCTCTAGCTCCGGATGTTGCTG 861

Qy 239 AsnAspPheLeuAsnAspGlnAsnArgAspValAsnSerIleLeuTyrSerGlyLeuGluAla 258
Db 862 AATGACTTCCTTAATGACCAGAATAGAGATAAAATAGTATCTATAGTGGCCTGGAAGCC 921

Qy 259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysAlaPheMet 278
Db 922 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 981

```
Qy  279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982 AAGGCTAGCTCTCGAAACCTTCAGTCAGAAAAACAAAAGGCATCTTTACTTGTATCAAATCAA 1041

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuValaSerValLeuAlaGlnTyr 318
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042 AATGGAAATATGTACACATCTTCAGTATATAGGTTCCCTTGCATCTGTTCTAGACACAGTAC 1101

Qy  319 SerProGlnLeuAlaGlyGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102 TCACCTCAGCAATTAGCAGGGGAGAGAATTGGAGTGTTTTCTTATGGGTTCTGGTTTGGCT 1161

Qy  339 AlaThrLeuTyrSerLeuLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162 GCCACTCTGTACTCTCTTAAAGTCACACAGGATGCTACACCGGGTTCTGCTCTTGATAAA 1221

Qy  359 IleThrAlaSerLeuCysAsnSerValTyrGlySerArgGlnThrGlyValAlaAlaPro 378
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222 ATAACAGCAAGTTTATGTAATTCAGTCTATGGTTCAAGAACAGGGGTTGGGGCACCA 1281

Qy  379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisIleLeuValAsnTyrIle 398
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTGGTCAACTATATT 1341

Qy  399 ProGlnGlySerIleAspSerLeuAspPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1342 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1401

Qy  419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1402 AAGCACAGAAGAACTTACGGCTCCGGCGTCCACTCCAAATGATGACACTTTGGATGAAGGA 1461

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrArgLeuHisIleProSerProAlaLysLysVal 458
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1462 GTAGGACTTGTGCATTCAAACATAGCAACTCGACTAATAATTCCAAGCCCCTGCCAAGAAAGTA 1521

Qy  459 ProArgLeuProAlaThrAlaAlaGluProGluLeuAlaAlaValIleSerAsnGlyGluHis 478
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1522 CCAAGACTTCCCTGCCACAGCAGCAGAACCTGAGCACTGTGCATTAGTAATGGGGGACAT 1581
```

```
RESULT 13
XX
AC  ACC62232;
XX
ID  ACC62232 standard; cDNA; 1650 BP.
XX
AC  ACC62232;
XX
DT  23-JUN-2003 (first entry)
XX
DE  Human NOV41g encoding cDNA SEQ ID NO:193.
XX
KW  Human; NOVX; antiatherosclerotic; hypotensive; cardiant; dermatological;
KW  anorectic; immunosuppressive; cytostatic; antidiabetic; antiinfertility;
KW  haemostatic; antiinflammatory; antiasthmatic; anti-HIV; immunomodulator;
KW  neuroprotective; nootropic; antiparkinsonian; metabolic; antilipaemic;
KW  gene therapy; cardiomyopathy; atherosclerosis; hypertension; scleroderma;
KW  congenital heart defect; aortic stenosis; valve disease; transplantation;
KW  tuberous sclerosis; obesity; congenital adrenal hyperplasia; diabetes;
KW  prostate cancer; metabolic disorder; neoplasm; lymphoma; uterus cancer;
KW  fertility; haemophilia; hypercoagulation; graft versus host disease;
KW  idiopathic thrombocytopenic purpura; AIDS; bronchial asthma; anorexia;
KW  Crohn's disease; multiple sclerosis; infectious disease; Parkinson's
KW  cancer-associated cachexia; Alzheimer's disease; Parkinson's disease;
KW  immune disorder; haematopoietic disorder; dyslipidaemia; cancer;
KW  metabolic syndrome X; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO2003023001-A2.
XX
PD  20-MAR-2003.
XX
PF  09-SEP-2002; 2002WO-US028538.
XX
PR  07-SEP-2001; 2001US-0318120P.
PR  07-SEP-2001; 2001US-0318184P.
PR  10-SEP-2001; 2001US-0318430P.
PR  17-SEP-2001; 2001US-0322636P.
PR  17-SEP-2001; 2001US-0322781P.
PR  17-SEP-2001; 2001US-0322816P.
PR  17-SEP-2001; 2001US-0322817P.
PR  19-SEP-2001; 2001US-0323519P.
PR  20-SEP-2001; 2001US-0323631P.
PR  20-SEP-2001; 2001US-0323636P.
PR  25-SEP-2001; 2001US-0324999P.
PR  26-SEP-2001; 2001US-0325091P.
PR  26-SEP-2001; 2001US-0325091P.
PR  14-DEC-2001; 2001US-0341144P.
PR  26-FEB-2002; 2002US-0359599P.
PR  05-MAR-2002; 2002US-0361663P.
PR  03-MAY-2002; 2002US-0377908P.
PR  17-MAY-2002; 2002US-0381483P.
PR  29-MAY-2002; 2002US-0383863P.
PR  02-JUL-2002; 2002US-0393332P.
PR  17-JUL-2002; 2002US-0396412P.
PR  13-AUG-2002; 2002US-0403517P.
PR  06-SEP-2002; 2002US-00226417.
XX
PA  (CURA-) CURAGEN CORP.
XX
PI  Agee ML, Alsobrook JP, Anderson DW, Berghs C, Boldog FL;
PI  Burgess CE, Casman SU, Catterton E, Chant JS, Chaudhuri A;
PI  Crabtree J, Dipippo VA, Edinger SR, Eisen AJ, Ellerman K;
PI  Gangolli EA, Gerlach VL, Giot L, Gorman L, Guo X, Guev VY, Ji W;
PI  Kekuda R, Khramtsov NV, Leach MD, Lepley DM, Li L, Liu X;
PI  Malyankar UM, Miller CE, Ooi CE, Ort T, Padigaru M, Patturajan M;
PI  Pena CEA, Rieger DK, Rothenberg ME, Shenoy SG, Shimkets RA;
PI  Spaderna SK, Spytek KA, Taupier RJ, Twomlow N, Vernet CAM, Voss EZ;
PI  Zerhusen BD, Zhong M;
XX
DR  WPI; 2003-313241/30.
DR  P-PSDB; ABR54463.
XX
PT  Novel human proteins and nucleic acid encoding the proteins, useful for
PT  diagnosis, treatment and prevention of disorders involving the human
PT  protein or nucleic acid e.g. cardiac and neurological disorders.
XX
PS  Claim 20; Page 243; 460pp; English.
XX
CC  The present invention describes isolated human NOVX proteins, where X is
CC  1 to 42. ACC62236 to ACC62345 encode the human NOVX proteins given in
CC  ABR54167 to ABR54276. NOVX sequences have antiatherosclerotic, cardiant,
CC  hypotensive, dermatological, anorectic, immunomodulator, immunosuppressive,
CC  antidiabetic, antiinfertility, haemostatic, antiinflammatory, anti-HIV,
CC  antiasthmatic, metabolic, anticancer, neuroprotective, nootropic,
CC  antiparkinsonian and antilipaemic activities. NOVX sequences can be used in gene
CC  therapy. NOVX proteins are useful for treating or preventing a pathology
CC  associated with a NOVX protein in humans and for treating a syndrome
CC  associated with the human disease. NOVX nucleic acids, proteins and
CC  antibodies can be used in the treatment and diagnosis of cardiomyopathy,
CC  atherosclerosis, hypertension, congenital heart defects, aortic stenosis,
CC  valve disease, tuberous sclerosis, scleroderma, obesity, transplantation,
CC  congenital adrenal hyperplasia, prostate cancer, diabetes, metabolic
CC  disorders, neoplasm, lymphoma, uterus cancer, fertility, haemophilia,
CC  hypercoagulation, idiopathic thrombocytopenic purpura, graft versus host
CC  disease, AIDS, bronchial asthma, Crohn's disease, multiple sclerosis,
CC  infectious disease, anorexia, cancer-associated cachexia, cancer,
CC  Alzheimer's disease, Parkinson's disease, immune disorders,
CC  haematopoietic disorders, dyslipidaemias, and metabolic syndrome X.
CC  ACC62346 to ACC62465 represent PCR primers and probes for human NOVX
CC  sequences, which are used in examples from the present invention.
CC  ABR54277 represents a human trypsinogen protein given in comparison with
CC  the human NOV35b protein in the exemplification of the present invention
XX
SQ  Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

Alignment Scores:

Pred. No.:	1.17e-249	Length:	1650
Score:	2480.00	Matches:	478
Percent Similarity:	91.92%	Conservative:	0
Percent Identity:	91.92%	Mismatches:	0
Best Local Similarity:	98.77%	Indels:	42
Query Match:	7	Gaps:	1
DB:			

US-10-622-516-2 (1-478) x ACC62332 (1-1650)

```
Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTTGGGATTGTT 81

Qy     21 AlaLeuGluIleTyrPheProSerGlnTyrValAlaAspGlnAlaGlnLeuGluLysTyrAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 GATAGAAGAAGAAATTAACTCTCTTTGCATGACTGGTTCAGAATCTTAYGAGGAGAAT 81

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyTyrLeuGlyAlnAlaLysMetGlyPheCysThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 GCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGACAGCAGAGTTGGAAAAATATGAT 141

Qy     61 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 GGTGTAGATGCTGGGAAGTATACACATTGGCTTGGGCTGGGCCGTTCTGCACA 201

Qy     81 AsnLeuSerTyrAspCysIleGlyTyrGluValValGlyThrGluThrIleTyrIle 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 AACCTTTCCTATGATTGCATTGGGCGGGGAAGTTGGAACAGAGACAAATCATCCGACAAA 261

Qy    101 GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyIleGlyIleThrAlaAlaValPheAsnAlaVal 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCGGGAATACAGATATA 381

Qy    121 AsnTrpIleGluSerSerSerTrpAsp 149
          ||||||||||||||||||||||||||
Db    442 AACTGGATTGAGTCCAGCTCTTGGGATGGA 501

Qy    149 ------------------------ 149
          
Db    502 GCTGTATATGCCACAGGAGAAATGCTAGACCTTACAGGTGGGAGTTGGAGCAGTAGCTCTGCTA 561

Qy    150 ----------------GlyLeuArgGlyThrHisMetGlnHis 158
                          |||||||||||||||||||
Db    562 ATTGGGCCAAAATGCTCCTTTAATTTTTGAACGAGGGGCTTCGTCGGACACATAAYGCAACAT 621

Qy    159 AlaTyrAspPheTyrArgLysProAspMetLeuSerGlyTyrProIleValAlaAspGlyLysLysLeu 178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622 GCCTATGATTTTTACAGAAGCCCTGATATGCTATCTGGATAYGATAGGGGAAAACTC 681

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAsnAspArgCysTyrSerValTyrCysLysLysIle 198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 TCCATACAAGTGCTACCTCAGTGCATTACCACCGCCTGCTATTCTGTGTCTACTGCAAAAAGATC 741

Qy    199 HisAlaGlnIleTrpGlnLysGlyGlyValAsnAspSerPheThrLeuAsnAspPheGlyPhe 218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742 CATGCCCAGTGGCAGAAAGGAGGAAATGAATAAAGATTTTACCTTGAAATGATTTTGGGCTTC 801

Qy    219 MetIlePheHisSerProTyrCysValGlnLysValGluAlaArgMetLeuLeu 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802 ATGATCTTTCACTCACCATATTGTAACTGGTTCAGAAAATCTTCAGCTCGGATGTTGCTG 861

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgAspGlyValAsnSerIleTyrSerGlyLeuGluAla 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAAAGTAGTATCCTAYAGTGGCCTGGAAGCC 921

Qy    259 PheGlyAspValLysGlyLeuGluAspThrTyrPheAspArgAspValGluLysValAlaPheMet 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    922 TTTGGGGATGTTAAATTAGGAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 981
```

```
PR   10-APR-2002; 2002US-0371507P.
PR   16-APR-2002; 2002US-0372984P.
PR   19-APR-2002; 2002US-0374194P.
PR   24-MAY-2002; 2002US-0382995P.
PR   31-MAY-2002; 2002US-0385042P.
PR   14-JUN-2002; 2002US-0388853P.
PR   17-JUN-2002; 2002US-0389395P.
PR   25-JUN-2002; 2002US-0391324P.
PR   15-JUL-2002; 2002US-0395944P.
PR   22-JUL-2002; 2002US-0397726P.
PR   13-AUG-2002; 2002US-0403046P.
PR   22-AUG-2002; 2002US-0405155P.
PR   27-AUG-2002; 2002US-0406361P.
PR   25-OCT-2002; 2002US-0421195P.
PR   12-NOV-2002; 2002US-0425456P.
PR   19-NOV-2002; 2002US-0427626P.
PR   10-DEC-2002; 2002US-0432122P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Hunter JJ, Macbeth KJ, Tsai F, Lesoon A, Lightcap ES;
PI   Williamson MW, Rudolph-Owen LA;
XX
DR   WPI; 2003-646176/61.
DR   P-PSDB; ADE38395.
XX
PT   Treating subject having tumorigenic disorder or angiogenic disorder
PT   caused by aberrant polypeptide e.g., N-formylpeptide receptor or nucleic
PT   acid, by administering a modulator.
XX
PS   Disclosure; SEQ ID NO 55; 454pp; English.
XX
CC   This invention relates to a novel method of treating a human subject
CC   having a tumourigenic disorder or angiogenic disorder, caused by aberrant
CC   gene expression or activity of an isolated protein, by administering a
CC   modulator. The modulator may have cytostatic, antithyroid, antidiabetic
CC   or ophthalmological activity. The method is useful for treating a subject
CC   having a tumourigenic or angiogenic disorder, in particular for treating
CC   cancer (for example breast cancer, colon cancer, lung cancer or prostatic
CC   cancer) and, for example, Grave's disease and diabetic retinopathy. The
CC   present sequence is a DNA sequence which encodes the novel isolated human
CC   protein 9389 of the invention.
XX
SQ   Sequence 1650 BP; 477 A; 319 C; 397 G; 457 T; 0 U; 0 Other;
```

```
Alignment Scores:

Pred. No.:
Pred:                      1.17e-249
Score:                     2480.00
Percent Similarity:        91.92%
Best Local Similarity:     91.92%
Query Match:               98.77%
DB:                        9

Length:         1650
Matches:        478
Conservative:   0
Mismatches:     0
Indels:         42
Gaps:           1

US-10-622-516-2 (1-478) x ADE38394 (1-1650)
```

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 ATGCCTGGAATCACTCCTTTGAATGCAGAAGCTTGCTCGGCCAAAGATGTTGGAATTGTT 81

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GCCCTPGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 141

Qy   41 GlyValAspAlaGlyGlyLysTyrThrIleGlyLeuGlyAsnLeuMetGlyPheCysThr 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 GGTGTAGATGCTGGCGGAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 201

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 261

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 AACCTTTCCTTATGATTGATTGGGCGGGGCTGGAAGTTGGAGACAGAGAACAATCGCAAA 321

Qy  101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAsnIle 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 TCAAAGTCTCGTGAAGACTAATTTGATGCAACTGTTTGTTGAAGAGTCTGGGAATACAGATATA 381

Qy  121 GluGlyIleIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT 441

Qy  141 AsnTrpIleGluSerSerSerTrpAsp---------------GlyLeuArgGlyTyrThrHisMetGlnHis 158
        |||||||||||||||||||||||||                |||||||||||||||
Db  442 AACTGGATTGAGTCCAGCTCTTGGGATGGACAGGTTGTCCCTGGAGTTGCAGGAGATATT 501

Qy  149 ----------- 149
Db  502 GCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGGAGTTGGAGCAGTAGCCTGCTA 561

Qy  159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspAspGlyLysLysLeu 178
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  562 ATTGGGCCAAATGCTCCTTTAATTTTTTGACGAGGGGCTTCGTGGGACACATATGCAACAT 621

Qy  179 SerIleGlnCysTyrLeuSerAlaLeuAspAlaArgCysTyrSerValTyrCysLysSerLysIle 198
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622 GCCTATGATTTTTACAAAGCCTGATATGCTAATCTGAATAATCCTATAGTAGATGGGAAACTC 681

Qy  199 HisIleAlaGlnHisTrpGlnLysSerGlyLysGlyAsnAspAspPheThrLeuAsnAspLysPheGlyPhe 218
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682 TCCATACAGTGCTACCTGAGTGCATTAGACCGGTGCTATTCTGTCTATTGTAAAAGATC 741

Qy  219 MetIlePheHisSerProTyrCysGlyLysLeuGluValGlnLysSerLeuValAlaArgMetLeuLeu 238
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742 CATGCCCAGTGCCAGAAAAGATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 801

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspLysSerAsnSerIleTyrSerGlyTyrLeuGluAla 258
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802 ATGATATTTCACTCACCATATGTAAAACTGGAAGTCAGAAATCTCAGCTCCGGATGTTGCTG 861

Qy  259 PheGlyAspValLysLysSerLeuValAspPheThrTyrPheAsnAspArgAspValGlyLysValAlaPheMet 278
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTAGAAGCC 921

Qy  279 LysAlaSerSerGluLeuPheSerGlnLeuLysThrLysAlaSerLeuLeuValSerAsnGln 298
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922 TTTGGGGATGTTAAATTAGAAGACACCTACTTTAATGATAGAGATGTGGAGAGGCATTTATG 981

Qy  299 AsnGlyAsnMetTyrThrSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982 AAGGCTAGCTCTGAACTCTTTCAGTCAAGGCAATCTTTACTGTGTATCAAATCAA 1041

Qy  319 SerProGlnGlnLeuAlaGlyIleGlyValPheSerTyrGlySerGlyLeuAla 338
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042 TCACCTCAGCAAATTGCAGGGGAAGAGAAATTGGAGTTTTCTTATGGTTCAGGGCT 1101

Qy  339 AlaThrLeuTyrSerLeuTyrThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102 GCCACTCTGTACTCTCTATAAAAGTCACACAAGACGCGGGTCTGCTCTTGATAAA 1161

Qy  359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162 ATAACAGCAAGTTTATGTGATCTTAAAATCAAGGCCTTGAATTCAAGGAACTGGTGCCACCA 1221

Qy  379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisSerLeuValAsnTyrIle 398
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222 GATGTCTTCGCTGAAAACATGAAAGCTCAGAGAGACACCCATCATTCTGTCAACTATATT 1281

Qy  399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGGTGGTACTTAGTGGGGTGGATGAA 1341
```

```
Qy      419  LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly  438
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1402 AAGCACAGAAGGAACTTTACCGCTCGGCGTCCCACTCCAAATGAATGAACACTTTGGATGATGAAGGA  1461

Qy      439  ValGlyLeuValHisSerAsnIleAlaThrGluHisHisSerAsnGlyGluHis  458
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1462 GTAGGGACTTGTGCATTCAAACATAGCCACTGAACATATTCCAAGGCCCTGCCAAGAAGTA  1521

Qy      459  ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis  478
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1522 CCAAGACTCCCTGCCACAGCAGCGAACCTGCTGTCATTAGTAATGGGAAACAT  1581
```

```
Score:                 2480.00         Matches:          478
Percent Similarity:    91.92%          Conservative:     0
Best Local Similarity: 91.92%          Mismatches:       0
Query Match:           98.77%          Indels:           42
DB:                    10              Gaps:             1

US-10-622-516-2 (1-478) x ADE76933 (1-3722)

Qy      1    MetProGlySerLeuProLeuAsnAlaAlaGluAlaCysTrpProLysGlyIleVal  20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      249  ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGGATGTGGGAATTGTT  308

Qy      21   AlaLeuGluIleTyrArgPheProSerGlnThrValValGlnAsnLeuMetGluArgAsn  40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      309  GCCCTTGAGATCTATTTTCCTTCTCAAATATGTTGATCAAAGCAGAGTTGGAAAATATGAT  368

Qy      41   GlyLeuAspLeuGlyLysTyrThrIleGlyLeuGlyGlnAlaLeuLysMetGlyPheCysThr  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      369  GGTCTAGATCTGGGAAAGTATACCATTGGTCTTGGGCAGGCCAAGAAATGGGGCTTCTGCACA  428

Qy      61   AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      429  GATAGAGAAGATATTAAACTCTCTTTGCATGACAGTGTGGTTCAAAATCTTATGGAGAGAAAT  488

Qy      81   AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGlyThrIleIleIleLeuLys  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      489  AACCTTTCCTATGATTGCATTGGCCGGCTGGAAGTTGGAACAGGAACAATCATCGACAAA  548

Qy      101  SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      549  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAAATACAGATATA  608

Qy      121  GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal  140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      609  GAAGGAATCGACACAACTAATGCATATGGTGGCACAGCTGCTGTCTTCAATGCTGTT  668

Qy      141  AsnTrpIleGluSerSerSerTrpAsp---------------------  149
             ||||||||||||||||||||||||||
Db      669  AACTGGATTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTAGTTGCAGGAGATATT  728

Qy      149  ------------------------------------GlyLeuArgGlyThrHisMetGlnHis  158
                                                 |||||||||||||||||||
Db      729  GCTGTATATCCCACAGGAAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA  788

Qy      150  ------------------------------AttTrpGGGCCAAATGCTCCTTTAATTTTTGAACGAGAGGCTTCGTGGGACACATATGCAACAT  848
Db      789  ATTTGGGCCAAATGCTCCTTTAATTTTTGAACGAGAGGCTTCGTGGGACACATATGCAACAT  848

Qy      159  AlaTyrAspPheTyrLysProAspMetLeuSerGluThrTyrProIleValAspGlyTyrLeu  178
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      849  GCCTATGATTTTTACAAGCCTGATATGCTATCTGAAACTTATCCTATAGTAGATGGATACTTG  908

Qy      179  SerIleGlnCysTyrLeuSerAlaLeuAsnArgCysTyrSerValTyrCysLysLysLysIle  198
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      909  TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC  968

Qy      199  HisSerLeuAlaGlnThrTrpGlnLysGlyLyAsnAspSerAspPheThrLeuAsnAspPheGlyPhe  218
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      969  CATGCCCAGTGGCAGAAAGGGAAATGATTAAAGGATTTTACCTTGAATGATTTTGGCTTC  1028

Qy      219  MetIlePheHisSerProTyrCysLysProGlyLeuValGlnLysSerLeuValArgMetLeuLeu  238
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1029 ATGATCTTTCACTCACCATATTGTAAACCTGGTTTCAGAAATCTCTAGCTCCGAGTGTTGCTG  1088

Qy      239  AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1089 AATGACTTCCTTAATGACCAGAATAGAGACAAGAATAGTATCTATAGTGGCCTGGAAGCC  1148

Qy      259  PheGlyAspValLysLeuGluAspThrThrTyrPheAspArgAspValGlyLysAlaPheMet  278
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1149 TTTGGGGATGTTAAATTAGAAGACACCACTTACTTTGATAGAGATGTGGGAAAGGCATTTATG  1208

Qy      279  LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln  298
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1209
```

```
Db   1209 AAGGCTAGCTCTGAAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTTATCAAATCA 1268
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1269 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1328

Qy    319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1329 TCACCTCAGCAATTAGCAGGGGAAGAGAATTGGAGTGTTTTTCTTATGGGTTTCTGGTTTGGCT 1388

Qy    339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1389 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCCGGGTCTGCTCTTGATAAA 1448

Qy    359 IleThrAlaSerLeuCysAspLeuLysAspSerArgLeuAspSerArgThrGlyValAlaPro 378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1449 ATAACAGCAAGTTTATGTGATCTTAAAATCAAGGCTTGATTCAAGAACTGGTGTCGGCACCA 1508

Qy    379 AspValPheAlaGluLeuAsnMetLysLeuArgGluAspThrHisHisLeuLeuValAsnTyrIle 398
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1509 GATGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAACTATATT 1568

Qy    399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1569 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1628

Qy    419 LysHisArgArgThrTyrAlaArgProThrProAsnAspAspThrLeuLeuAspGlyGly 438
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1629 AAGCACAGGAGAACTTACCGTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGA 1688

Qy    439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1689 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCCATTATTCCAAGCCCTGCCCAAGAAAGTA 1748

Qy    459 ProArgLeuProAlaThrAlaAlaGluProGluAlaLeuAlaValIleSerAsnGlyGluHis 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1749 CCAAGACTCCCTGCCACACGAGCAGAACCTGAAGCAGCTGCTGTCATTAGTAATGGGGAACAT 1808
```

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 15:10:46 ; Search time 524 Seconds
                                    4178.933 Million cell updates/sec

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV............PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext 7.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1p
 -Q=/cgn2_1/USPTO_spool_p/US10622516//runat_23062004_162636_775/app_query.fasta_1.647
 -DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
 -LOOPCL=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -HEAPSIZE=500 -MINLEN=0
 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -THR_MAX=100
 -MAXLEN=200000000 -USER=US10622516 @CGN_1_1_723 @runat_23062004_162636_775
 -NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY_NEG_SCORES=0 -WAIT -XGAPOP=10
 -LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -DSPBLOCK=100
 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Published_Applications_NA:*
   1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
   2: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
   3: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
   4: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
   5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
   6: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
   7: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
   8: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
   9: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
  10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
  11: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
  12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
  13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
  14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
  15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
  16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
  17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
  18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
  19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Score	Query Match Length	DB	ID	Description
1	2511	100.0	2002	14	Sequence 1, Appli
2	2511	100.0	2002	16	Sequence 1, Appli
3	2480	98.8	1564	13	Sequence 203, App
4	2480	98.8	1564	13	Sequence 207, App
5	2480	98.8	1564	13	Sequence 369, App
6	2480	98.8	1564	13	Sequence 373, App
7	2480	98.8	1601	13	Sequence 191, App
8	2480	98.8	1601	13	Sequence 363, App
9	2480	98.8	1601	13	Sequence 181, App
10	2480	98.8	1650	13	Sequence 183, App
11	2480	98.8	1650	13	Sequence 185, App
12	2480	98.8	1650	13	Sequence 187, App
13	2480	98.8	1650	13	Sequence 193, App
14	2480	98.8	1650	13	Sequence 197, App
15	2480	98.8	1650	13	Sequence 201, App
16	2480	98.8	1650	13	Sequence 205, App
17	2480	98.8	1650	13	Sequence 359, App
18	2480	98.8	1650	13	Sequence 371, App
19	2480	98.8	1650	13	Sequence 445, App
20	2480	98.8	1650	15	Sequence 55, Appl
21	2480	98.8	3722	10	Sequence 98, Appl
22	2480	98.6	1593	13	Sequence 187, App
23	2475	98.6	1593	13	Sequence 361, App
24	2475	98.6	1608	13	Sequence 195, App
25	2475	98.6	1608	13	Sequence 211, App
26	2475	98.6	1608	13	Sequence 365, App
27	2475	98.6	1608	13	Sequence 375, App
28	2475	98.6	1612	13	Sequence 199, App
29	2475	98.6	1612	13	Sequence 209, App
30	2475	98.6	1612	13	Sequence 377, App
31	2475	98.6	1612	13	Sequence 367, App
32	2362	94.1	3275	13	Sequence 15, Appl
33	2362	94.1	3275	16	Sequence 37, Appl
34	1691	67.3	170834	9	Sequence 7, Appli
35	1691	67.3	170834	15	Sequence 7, Appli
36	1631	65.0	3008	13	Sequence 1926, Ap
37	1513	60.3	2058	9	Sequence 417, App
38	1513	60.3	2058	9	Sequence 3810, Ap
39	1513	60.3	2058	13	Sequence 1557, Ap
40	1513	60.3	2058	13	Sequence 1161, Ap
41	1513	60.3	2058	13	Sequence 1161, Ap
42	1513	60.3	2058	13	Sequence 173, App
43	1513	60.3	2058	13	Sequence 207, App
44	1513	60.3	2261	15	Sequence 10322, A
45	1513	60.3	2377	15	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-10-193-295-1
Sequence 1, Application US/10193295
Publication No. US20020173018A1
GENERAL INFORMATION:
  APPLICANT: GONG, Fangcheng et al.
  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
  TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
  TITLE OF INVENTION: THEREOF
  FILE REFERENCE: CL001195DIV
  CURRENT APPLICATION NUMBER: US/10/193,295
  CURRENT FILING DATE: 2002-07-12
  PRIOR APPLICATION NUMBER: 08/819,993
  PRIOR FILING DATE: 2001-03-29
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 1
  LENGTH: 2002
  TYPE: DNA
  ORGANISM: Human
  US-10-193-295-1

Alignment Scores:

```
Pred. No.:
Score:                    2.39e-293
Percent Similarity:       2511.00          Length:          2002
Best Local Similarity:    100.00%          Matches:          478
Query Match:              100.00%          Conservative:       0
DB:                       100.00%          Mismatches:         0
                          14               Indels:             0
                                           Gaps:               0
```

US-10-622-516-2 (1-478) x US-10-193-295-1 (1-2002)

```
Qy      1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal  20
Db    145  ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 204

Qy     21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluGlyLysTyrAsp  40
Db    205  GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGCTTGAAAAATATGAT 264

Qy     41  GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyAlaLysMetGlyPheCysThr  60
Db    265  GGTGTAGAGATGGCTGGAAAGTATACCATTGGCCTTGGGCAAAGATGGGCTTCTGCACA 324

Qy     61  AspProArgGluAsnAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn  80
Db    325  GATAGAGAGAGATAATTAACTCTTTGCATGACTGTTCAGAATCTTATGGAGAGAAAT 384
```

...

; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00119501V2
; CURRENT APPLICATION NUMBER: US/10/622,516
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 10/193,295
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-10-622-516-1

Alignment Scores:
```
Pred. No.:
Score:                    2.39e-293
Percent Similarity:       2511.00          Length:          2002
Best Local Similarity:    100.00%          Matches:          478
Query Match:              100.00%          Conservative:       0
DB:                       100.00%          Mismatches:         0
                          16               Indels:             0
                                           Gaps:               0
```

US-10-622-516-2 (1-478) x US-10-622-516-1 (1-2002)

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaIleLysMetGlyGlyPheCysThr 60
Db   265 GGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCAGGCCAAAGATGGGCTTCTGCACA 324

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
Db   325 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 384

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGlyValGlyThrGluThrIleIleIleAspLys 100
Db   385 AACCTTTCCTATGATTGCATTGGGCGCCTGGGAGTTGGAACAGAGACAAATCATCGACAAA 444

Qy   101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIle 120
Db   445 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAGAGTGGGAATACAGATATA 504

Qy   121 GluGlyIleAspThrThrAsnAlaCysSerTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
Db   505 GAAGGAATCGACACAACTAATGCATGCTCTTATGGGGGAACTGCTGCTGTTTTCAATGCTGTT 564

Qy   141 AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyTyrThrHisMetGlnHisAlaTyr 160
Db   565 AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGGCTACACTCATATGCAACATGCCTAT 624

Qy   161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLysLeuSerIle 180
Db   625 GATTTTTACAAGCCTGATATGCTTAGTGAATATCCTATAGTAGATGGAAAACTCTCATA 684

Qy   181 GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysSerGlyLysHisAla 200
Db   685 CAGTGCTACCTCAGTGCATTAGACCGGTGCTACTCTGTCTACTGCAGTGGAAAAGATCCATGCC 744

Qy   201 GlnTrpGlnLysGlyLysGlyAsnAspPheThrLeuAsnAspPhePheGlyPheMetIle 220
Db   745 CAGTGGCAGAAAGGGAAATGATAAAGATTTTTACCTTGAATGATTTTTGGCTTCATGATC 804

Qy   221 PheHisSerProTyrCysLysLysGluValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Db   805 TTTCACTCACCATATTGTAAACTGGGTTCAGAAATCTCTAGCTCGGAGTCCTCGGAATGAC 864

Qy   241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
Db   865 TTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCCTGGAAGCCTTTGGG 924

Qy   261 AspValLysGluAspThrThrTyrPheAspArgAspArgValGlyLysAlaPheMetLysAla 280
Db   925 GATGTTAAATTAGAAGACACCCTACTTTGATAGGAGAAGGCATTTATCAAAGGCT 984

Qy   281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGly 300
Db   985 AGCTCTGAACTCTTCAGTCAGAAAACAAAAGGCATCTTTACTTGTGTATCAAAATCAAAATGGA 1044

Qy   301 AsnMetTyrThrSerValTyrGlySerLeuGluAlaGlnInTyrSerPro 320
Db 1045 AATATGTACACATCTTCAGTATATGGTTCCCTTGCAATCTGTTCTAGCACAGTACTCACCT 1104

Qy   321 GlnGlnLeuAlaGlyIleGlyValPheSerGlyValLeuGlyLeuAlaAlaThr 340
Db 1105 CAGCAAATTAGCAGGGAAGGAATTGGAGTGTTTTCTTATGGTTCTCGGGTTCTGGGTGCCACT 1164

Qy   341 LeuTyrSerProLeuGlyLysValThrArgGlnAspAlaAlaThrProGlySerAlaAlaLeuAspLysIleIleThr 360
Db 1165 CTGTACTCTCCTAAAGTCACAAGATGCTACACGGGGGTCTCGTCTCTGATAAAATAACA 1224

Qy   361 AlaSerLeuCysAspLeuLysSerArgLeuAsnAspSerArgThrGlyValAlaProAspVal 380
Db 1225 GCAAGTTTATGTGATCTTAAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCAGATGTC 1284

Qy   381 PheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAlaAsnTyrIleProGln 400
Db 1285 TTCGCTGAAAACATGAAGCTCAGAGAAGACACCCATCATTTGGTCAACTATATTCCCCAG 1344

Qy   401 GlySerIleAspSerLeuPheGluGlyThrThrpTyrLeuValAlaArgValAspGlyLysHis 420
Db 1345 GGTTCAATAGATTCACTCTTTGAGGGAACCACGTGGTACTTAGTTAGGGTCGATGGAAAGCAC 1404

Qy   421 ArgArgThrTyrAlaLeuArgArgProThrProAsnAspAspAspThrLeuAspGlyValValGly 440
Db 1405 AGAAGAACTTACGGTCGGCGGCCTCCCACTCCAAAATGACACTTTGGATGGAGTAGTAGGA 1464

Qy   441 LeuValHisSerAsnIleAlaThrGlnHisHisIleProSerProAlaLysLysValProArg 460
Db 1465 CTTGTGCATTCAAACATAGCAACTCAGCACATATTCCAAGCCCTGCCAAAGAAAGTACCAAGA 1524

Qy   461 LeuProAlaThrAlaAlaLeuGluProGluValAlaValIleIleSerAsnGlyGluHis 478
Db 1525 CTCCCTGCCACAGCAGCAGCACCTGAAGCAGAACCTGAAGTCGTCATTAGTAATGGGGAACAT 1578

Alignment Scores:
Pred. No.:
Score: 9.17e-290   Length: 1564
Percent Similarity: 2480.00   Matches: 1564
Best Local Similarity: 91.52%   Conservative: 0
Query Match: 91.92%   Mismatches: 0
DB: 98.77%   Indels: 42
                      13   Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-203 (1-1564)

Qy   1 MetProGlySerLeuProLeuAsnAlaGluLeuAlaCysLysTrpProLysGlyAspValGlyIleVal 20
Db   2 ATGCCTGGATCACTTCCTTTGAATGCAGAGCTTGCTGGGCCAAAAGGATGTGGGAATTGTT 61

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluGlyLysTyrAsp 40

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       62 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 121

Qy       41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      122 GGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 181

Qy       61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      182 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 241

Qy       81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      242 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAA 301

Qy      101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      302 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 361

Qy      121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      362 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT 421

Qy      141 AsnTrpIleGluSerSerSerTrpAsp--------------------------------- 149
            |||||||||||||||||||||||||||
Db      422 AACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTGCAGGAGATATT 481

Qy      149 ------------------------------------------------------------ 149

Db      482 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 541

Qy      150 ---------------------------------GlyLeuArgGlyThrHisMetGlnHis 158
                                             |||||||||||||||||||||||||||
Db      542 ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 601

Qy      159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      602 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC 661

Qy      179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      662 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATC 721

Qy      199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      722 CATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 781

Qy      219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      782 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 841

Qy      239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      842 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC 901

Qy      259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      902 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 961

Qy      279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      962 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAA 1021

Qy      299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1022 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1081

Qy      319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1082 TCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTTCTGGTTTGGCT 1141

Qy      339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1142 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCTCTTGATAAA 1201

Qy      359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1202 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA 1261

Qy      379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1262 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAACTATATT 1321

Qy      399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1322 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1381

Qy      419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1382 AAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGA 1441

Qy      439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1442 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1501

Qy      459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1502 CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT 1561
```

```
RESULT 4
US-10-236-417-207
; Sequence 207, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-442C
;  CURRENT APPLICATION NUMBER: US/10/236,417
;  CURRENT FILING DATE:  2003-01-06
;  PRIOR APPLICATION NUMBER: US60/318,120
;  PRIOR FILING DATE: 2001-09-01
;  PRIOR APPLICATION NUMBER: US60/318,430
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: US60/322,781
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/318,184
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US60/361,663
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US60/396,412
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: US60/322,636
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,817
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,816
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/323,519
;  PRIOR FILING DATE: 2001-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom
; SEQ ID NO 207
;   LENGTH: 1564
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2)..(1562)
US-10-236-417-207

Alignment Scores:
Pred. No.:              9.17e-290      Length:         1564
Score:                  2480.00        Matches:        478
Percent Similarity:     91.92%         Conservative:   0
Best Local Similarity:  91.92%         Mismatches:     0
```

Query Match: 98.77%   Indels: 42
DB: 13   Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-207 (1-1564)

```
Qy   1  MetProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProLysAspValGlyIleVal  20
Db   2  ATGCCTGGATCACTTCCTTTGAATGCAGGAGCTTGCTGGCCAAAAGATGGGAATTGTT  61

Qy  21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlnLysTyrAsp  40
Db  62  GCCCTCGAGAATCTATTTCCTTCTCAATATGTGATCAAGCAGAGTTGGAAAAAATATGAT  121

Qy  41  GlyValAspAlaGlyLysTyrThrIleGlyLeuGlnHisAlaLysMetGlyPheCysThr  60
Db 122  GGTGTAGATGCTGGGAAAGTATACCATTGGCTTGGGCAGCACGCCAAGATGGGGCTTTGCACA  181

Qy  61  AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn  80
Db 182  GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT  241

Qy  81  AsnLeuSerTyrAspCysTyrIleGlyArgLeuValGlyValGlyThrGluThrIleIle  100
Db 242  AACCTTTCCTATGATTGCTATATAGGGCGGCTTGGAAGTTGGAACAGAGACAAATCATCGAACAA  301

Qy 101  SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIle  120
Db 302  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAAGTGGAAATACAGATATA  361

Qy 121  GluGlyIleAspThrThrAsnAlaCysSerTyrGlyGlyThrAlaAlaValPheAsnAlaVal  140
Db 362  GAAGGAATCGACACAACTAATGCATGCTCTTATGGAGGAACAGCTGCTGTCTTCAATGCTGTT  421

Qy 141  AsnTrpIleGluSerSerSerTrpAsp-----------------------  149
Db 422  AACTGGATTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTAGTTGCAGGAGATATT  481

Qy 149  -------------------------------------  149
Db 482  GCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTGGAGCAGTAGCTCTGCTA  541

Qy 150  -------GlyLeuArgGlyThrHisMetGlnHis  158
Db 542  ATGGGGCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTGGGACACATATGCCAACAT  601

Qy 159  AlaTyrAspPheThrTyrLysProAspMetLeuSerGlyTyrProIleValAspGlyLeu  178
Db 602  GCCTATGATTTTACAAGCCTGATATGCTATCCTAATAGTACAGTAGATGGAAAACTC  661

Qy 179  SerIleGlyTyrLeuSerAlaLeuLeuAspArgCysTyrSerValTyrCysLysLysIle  198
Db 662  TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTTTACTGCAAAAAGATC  721

Qy 199  HisAlaGlnTrpProGlnLysGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe  218
Db 722  CATGCCCAGTGGCCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC  781

Qy 219  MetIlePheHisSerProThrCysTyrCysGlyLysValGlnLysSerLeuAlaArgMetLeuLeu  238
Db 782  ATGATCTTTCACTCACCATATTGTTACTGCGTTCAGAAATCTCTAGCTGGATGTTGCTG  841

Qy 239  AsnAspPheLeuAsnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
Db 842  AATGACTTCCTTAATGACAGAGATAAAAATAGTATCTATAGTGGCCCTGAAGCC  901

Qy 259  PheGlyAspValGlyAspArgAspAlaValGlyTyrAlaPheMet  278
Db 902  TTTGGGGATGTTAAATTAGGAGACACCTTACTTGGATAGAGAAGGCATTTATG  961

Qy 279  LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln  298
Db 962  AAGGCTAGTTCTGAACTCTTCAGTCAGAAAACAAAGGCTTCACTCTTTACTGTATCAAATCAA  1021
```

US-10-622-516-2 (1-478) x US-10-307-817-369 (1-1564)

```
Qy   1  MetProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProLysAspValGlyIleVal  20
Db   2  ATGCCTGGATCACTTCCTTTGAATGCAGGAGCTTGCTGGCCAAAAGATGGGAATTGTT  61

Qy  21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlnLysTyrAsp  40
```

RESULT 5
US-10-307-817-369
; Sequence 369, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-502C
; CURRENT APPLICATION NUMBER: US/10/307,817
; CURRENT FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 682
; SOFTWARE: CuraSeqlist version 0.1
; SEQ ID NO 369
; LENGTH: 1564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1561)
US-10-307-817-369

Alignment Scores:
Pred. No.:
Score: 9.17e-290   Length: 1564
Percent Similarity: 2480.00   Matches: 478
Best Local Similarity: 91.92%   Conservative: 0
Query Match: 91.92%   Mismatches: 0
DB: 98.77%   Indels: 42
          13   Gaps: 1

```
Qy 299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerGlySerValLeuAlaSerLeuAlaGlnThr  318
Db 1022 AATGGAAATATGTACACATCTTCAGTATATGGTTCCGGCAGTAGTAATTTGGCT  1081

Qy 319  SerProGlnGlnLeuAlaGlyLysArgIleGlyValProPheSerTyrGlySerGlyValLeuAla  338
Db 1082 TCACCTCAGCAAATTAGCAGGGAAGAGAAATTCAGGAGTTTTCTTATAGTGGTTTGGCT  1141

Qy 339  AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAlaAspLys  358
Db 1142 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTCTGCTTTGATAAA  1201

Qy 359  AspValPheAlaGluLeuAsnMetLysArgLeuTyrSerSerArgThrGlyValAlaPro  378
Db 1202 ATAACAGCAAGTTTAATGTGATCTTAAATCAAGGCTTGATCAAGGAACTGGTTGGTGTA  1261

Qy 379  IleThrAlaSerLeuCysTyrTrpLysLysSerArgLeuValAspSerValGlyValAlaPro  398
Db 1262 GATGTCTTCCCTGAAAAACATGAAGGCTTCAGAGAGGACACCCATCATTAGTTGCAACATTAT  1321

Qy 399  ProGlnGlySerLeuPheGluGlyThrTrpTyrLeuValArgValAlaArgValAspGlu  418
Db 1322 CCCCAGGGTTCAAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGGGTGGATGAA  1381

Qy 419  LysHisSerArgArgThrTyrArgLeuAlaArgProThrProAsnAspAspThrLeuAspSerGly  438
Db 1382 AAGCACAGAAGAACTTACGCTCGGCGGCCTCCAAATGATGACACTTTGGATGAAGGA  1441

Qy 439  ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysLysVal  458
Db 1442 GTAGGACTTGTGCATTCAAATATTGCAACTGAGCATATTCCAAGCCCTGCCAAGAAGAAGTA  1501

Qy 459  ProArgLeuProAlaThrArgAlaLysAlaGluProGlyAlaAlaValIleLeuSerAsnGlyGluHis  478
Db 1502 CCAAGACTCCCTGCCACAGAGGCGAAGGCAGAACCTGGAGCAGCTGTCATTAGTAATGGGGAACAT  1561
```

US-10-622-516-2 (1-478) x US-10-307-817-369 (1-1564)

```
Qy   1  MetProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProLysAspValGlyIleVal  20
Db   2  ATGCCTGGATCACTTCCTTTGAATGCAGGAGCTTGCTGGCCAAAAGATGGGAATTGTT  61

Qy  21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlnLysTyrAsp  40
```

```
Db         62 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 121

Qy         41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 181

Qy         61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 241

Qy         81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAA 301

Qy        101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 361

Qy        121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GAAGGAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTCTTCAATGCTGTT 421

Qy        141 AsnTrpIleGluSerSerSerTrpAsp--------------------------------- 149
              |||||||||||||||||||||||||||
Db        422 AACTGGATTGAGTCCAGCTCTTGGGATGGACGGTATGCCCTGGTAGTTGCAGGAGATATT 481

Qy        149 ------------------------------------------------------------ 149

Db        482 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 541

Qy        150 ---------------------------------GlyLeuArgGlyThrHisMetGlnHis 158
                                               |||||||||||||||||||||||||||
Db        542 ATTGGGCCAAATGCTCCTTTAATTTTTGAACGAGGGCTTCGTGGGACACATATGCAACAT 601

Qy        159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC 661

Qy        179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATC 721

Qy        199 HisAlaGlnTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 CATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 781

Qy        219 MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTCTAGCTCGGATGTTGCTG 841

Qy        239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC 901

Qy        259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet 278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTGGAGAAGGCATTTATG 961

Qy        279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAA 1021

Qy        299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1081

Qy        319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 TCACCTCAGCAATTAGCAGGGAAGAGAATTGGAGTGTTTTCTTATGGTTCTGGTTTGGCT 1141

Qy        339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1142 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCTCTTGATAAA 1201

Qy        359 IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaPro 378
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 ATAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA 1261

Qy        379 AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1262 GATGTCTTCGCTGAAAACATGAAGCTCAGAGAGGACACCCATCATTTGGTCAACTATATT 1321

Qy        399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1322 CCCCAGGGTTCAATAGATTCACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1381

Qy        419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1382 AAGCACAGAAGAACTTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGA 1441

Qy        439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1442 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1501

Qy        459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1502 CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT 1561

RESULT 6
US-10-307-817-373
; Sequence 373, Application US/10307817
; Publication No. US20040058338A1
; GENERAL INFORMATION:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-502C
;  CURRENT APPLICATION NUMBER: US/10/307,817
;  CURRENT FILING DATE:  2002-12-02
;  NUMBER OF SEQ ID NOS: 682
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 373
;   LENGTH: 1564
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2)..(1561)
US-10-307-817-373

Alignment Scores:
Pred. No.:                9.17e-290      Length:         1564
Score:                    2480.00        Matches:        478
Percent Similarity:       91.92%         Conservative:   0
Best Local Similarity:    91.92%         Mismatches:     0
Query Match:              98.77%         Indels:         42
DB:                       13             Gaps:           1

US-10-622-516-2 (1-478) x US-10-307-817-373 (1-1564)

Qy          1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 61

Qy         21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 121

Qy         41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 GGTGTAGATGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 181

Qy         61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 241

Qy         81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
```

RESULT 7
US-10-236-417-191
; Sequence 191, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
SEQ ID NO 191
; LENGTH: 1601
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1588)
US-10-236-417-191

Alignment Scores:
Pred. No.:
Score: 2480.00
                         9.52e-290
Percent Similarity: 91.92%   Length: 1601
Best Local Similarity: 91.92%   Matches: 478
                         98.77%   Conservative: 0
Query Match: 98.77%   Mismatches: 0
DB: 13   Indels: 42
                                      Gaps: 1

US-10-622-516-2 (1-478) x US-10-236-417-191 (1-1601)

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 72

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 132

Qy   141 AsnTrpIleGluSerSerSerTrpAsp--------------------------------- 149
         |||||||||||||||||||||||||
Db   422 AACTGGATTGAGTCCAGTTCTTGGGATGGACGGTATGCCCTGGGTAGTTGCAGGAGATATT 481

Qy   150 ------------------GlyLeuArgGlyThrHisMetGlnHis   158
                           |||||||||||||||||||||||
Db   482 GCTGTATATGCCACAGGGAAATGCTAGACCTACAGGGAGTTGGAGCAGTAGCTCTGCTA 541

Qy   159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 ATTGGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCTTCGTCGGGACACATATGCAACAT 601

Qy   179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerSerValTyrCysLysLysIle 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   602 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC 661

Qy   199 HisAlaGlnIleTrpGlnLysGlyGluGlyAsnAspProLysSAspPheThrLeuAsnAspPheGlyPhe 218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   662 TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCGTCTACTGCAAAAAGATC 721

Qy   239 AsnAspPheLeuAsnAspArgLysAsnSerIleTyrSerArgGlyLeuGluAla 258
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   722 CATGCCCAGTGGCAGAAAGGAGGAAAATGATAAAGATTTTACTTGATCAATGATTTTGGCTTC 781

Qy   259 PheGlyAspValLysLeuGluAspThrTyrPheAspAsnArgAspValGlyLysAlaPheMet 278
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   782 ATGATCTTTCACTCACCACATATTGTAAACTGGTTCAGAAATCCTGAGTTGGATGTTGCTG 841

Qy   279 LysSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   842 AATGACTTCCTTAATGACCAGAATAGAGAATAAAATAGTATCTTAAGTGGCCTGAAGCC 901

Qy   299 AsnGlyAlaAsnMetTyrThrSerSerValTyrGlySerLeuValLeuAlaGlnTyr 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   902 TTTGGGGATGTTAAAATTAGAAGACACACCTACTTTGATACGAGATGTGGGAAGGCATTTATG 961

Qy   319 SerProGlnGlnLeuAlaGlyLysGlyValPheSerTyrGlySerGlyLeuAla 338
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   962 AAGGCTAGCTTCTGAAACTCTTCAGTCAGAAAACAAAGGCATCTTACTTGTATCAAAATCAA 1021

Qy   359 IleThrAlaSerLeuCysAlaSerProLeuLysSerArgLysThrGlyValAlaPro 378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1022 AATGGAAATATGTACAACTTCAGTATATANGGTTCCCCTTGCATCTGTTCTAGACAGTAC 1081

Qy   379 AspValPheAlaGluAsnMetLysLeuArgGluThrHisHisLeuValLeuAsnTyrIle 398
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1082 TCACCCTCAGCAATTAGCAGGGAGGAAGAATTGGAGTGTTTTCTGGTTCTGGTTTGGCT 1141

Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValAlaArgValAspGlu 418
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1142 GCCACTCCTGTACTCTCTTAAGTCACAAGAGCTACACCGGGGTCCGCTCCTTGATPAA 1201

Qy   422 ... 481

Db   1202 ATAACAGGAAGTTTATGTGATCTTAAATCAAGCCTTGATTCAAGAAACTGGTCGTGTGCCACCA 1261

Db   1262 GATGTCTTCGCTGAAAAACATGAAGGCTCAGAGGAACCACCCATCATTTGGTCACTATATT 1321

Db   1322 CCCCAGGGGTTCAATAGAATTCACTCTTTTGAAGGAACGTGTGACTTAGTTACGGGTGGATGAA 1381

Qy   419 LysHisSerArgGlyThrTyrArgLeuArgArgProThrProAsnAspAspThrLeuAsnGluGlyGly 438
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1382 AAGCACAGAAGGAACTTACCGCTCGCGTCCCACTCCAAATGAATGACACTTTGGATGAAGGA 1441

Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1442 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1501

Qy   459 ProArgLeuProAlaThrAlaAlaLeuGluProGluAlaAlaLeuAlaVallleSerAsnGlyGluHis 478
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   1502 CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCGTGTCATTAGTAATGGGGAAACAT 1561

```
Qy    41  GlyValAspAlaGlyLysThrTyrThrIleGlyLeuGlyValGlnAlaAlaLysMetGlyPheCysThr      60
Db   133  GGTGTAGAtGtCTGGAAGaccTAtACcAttgGTcTtGGcGTTGGGCCaagGaTGGGGCttCtGCaCA          192

Qy    61  AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn                   80
Db   193  GATAGAGAAGAtAtTAAcTCtCTtTGcAtGACTGttGTtCAGAAtCttATGGAGAGAaat                   252

Qy    81  AsnLeuSerTyrArgAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys                100
Db   253  AAcCTtTCcTAtCgAttGCAtTGGGcGcGGCtGGAAGttGGAACaGAAGACAAtCAtCGaCAA               312

Qy   101  SerLysSerValLysThrAsnLeuMetGluLeuPheGluGlySerGlyAsnThrArgIle                   120
Db   313  TCAAaGTCTGtGAAgACtAAtTTGAtGGaGctGTTtGAgGGAtCtGGaAAtACaAGAATA                   372

Qy   121  GluGlyIleLeuAspThrThrAsnAlaCysTyrCysGlyIleThrAlaAlaValPheAsnAlaVal             140
Db   373  GAAGGAAtTGACACAACtAAtGCAtGCTAtTGtGgAGGCACaGCtGCTGTtTTCAAtGCTGtT               432

Qy   141  AsnTrpIleGluLeuSerSerSerThrpAsp                                                149
Db   433  AACtGGATtGAgTTgTCCAGcTCTTTgGGAtGacGGTATGGCCCTGTGAGTTGCAGGAGatATt             492
```

```
Alignment Scores:
Pred. No.:
Score:                  9.52e-290
Percent Similarity:     2480.00          Length:        1601
Best Local Similarity:  91.92%           Matches:       478
Query Match:            91.92%           Conservative:  0
                        98.77%           Mismatches:    0
                        13               Indels:        42
DB:                                      Gaps:          1
```

US-10-622-516-2 (1-478) x US-10-307-817-363 (1-1601)

```
Qy     1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleLeuVal               20
Db     1  ATGCCtGGAtCACTtCCtTTGAAtGCaGAAGCtTGCTGGCCAAAAGAtGTGGGAAtTGTT                  72
```

RESULT 9
US-10-236-417-181
; Sequence 181, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 181
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
US-10-236-417-181

Alignment Scores:
Pred. No.:
Score: 1e-289
Percent Similarity: 2480.00
Best Local Similarity: 91.92%
Query Match: 91.92%
DB: 98.77%

Length:	1650
Matches:	478
Mismatches:	0
Conservative:	0
Indels:	42
Gaps:	1

US-10-622-516-2 (1-478) x US-10-236-417-181 (1-1650)

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysValGlyIleVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGTAGGGATTGTT 81

Qy   21 AlaLeuGlyIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 141

```
Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyIleGlyIleGlnAlaLysMetGlyPheCysThr 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 GGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCATCCAGGCCAAGATGGGCTTCTGCACA 201

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGGAAT 261

Qy    81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleGluAspLys 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGAACAA 321

Qy   101 SerLysSerValLysThrAsnLeuPheGluSerGlyAsnThrAsnIle 120
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 TCAAAGTCTGTCAAGACTAATTTGTTTGAGAGTCGGAAATACAGATATA 381

Qy   121 GluGlyIleAspThrArgAsnAlaCysTyrGlyTyrThrAlaAlaValAlaValPheAsnAlaVal 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382 GAAGGAATCGACACAACTAAGCAATGCTATGGGGACACAGCTGCTGTCTTCAATGCTGTT 441

Db   141 AsnTrpIleGluSerSerSerTrpAsp----------------------- 149
         |||||||||||||||||||||||||||
Qy   442 AACTGGATTGAGTCCAGCTCTTGGGATGGACGGGTATGCCCTGGTAGTTGCAGGAGATATT 501

Qy   150 ------------------GlyLeuArgGlyTyrThrHisMetGlnHis 158
                           ||||||||||||||||||||||||||||
Db   502 GCTGTATATGCCACAGGAAATGCTAGACCTTACAGGTGGAGTTGGAGCAGTAGCTCTGCTA 561

Qy   159 AlaTyrAspPheTyrIleLysProcAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562 ATTGGGCCAAATGCTCCTTTAATTTTCGTCGGTGGACACATATGCAACAT 621

Qy   179 SerIleGluGlnCysTyrIleLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAACTC 681

Qy   199 HisAlaGlnThrTrpGlnLysGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   682 TCCATACAGTGCTACCTCAGTGCATTAGACCGGTGCTATTCTGTCTACTGCAAAAAGATC 741

Qy   219 MetIlePhePheHisSerProTyrCysTyrLysLysSerLeuAlaArgMetLeuLeu 238
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   742 CATGCCCAGTGGCAGAAAGGGGAAGATAAAGATTTTACCTTGAATGATTTTGGCTTC 801

Qy   239 MetIlePhePheHisSerProTyrCysTyrLysLysSerLeuAlaArgMetLeuLeu 258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   802 ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAAATCCTAGCTCGGATGTTGCTG 861

Qy   259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAsnValGlyLysAlaPheMet 278
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862 AATGACTTCCTTAATGACCAGAATAAAGATAGTTACTTATAGTGGCCTGGAGCC 921

Qy   279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   922 TTTGGGGATGTTAAATTGAAGACCTACTTGATAGAGAGGCCATTTATG 981

Qy   299 AsnGlyAsnMetTyrTyrThrSerValTyrGlySerLeuAlaLeuAlaGlnTyr 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982 AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAAATCAA 1041

Qy   319 SerProGlnGlnLeuLeuAlaGlyLysArgIleGlyValPheSerTyrSerGlyLeuAla 338
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1042 AATGGAAATATGTACACCTCAGTATATGGTTCCCTTGCATCTGTTGCACACAGTAC 1101

Qy   339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1102 TCACCTCAGCCAATTAGCACGGAGGAGTGGTTTTCTTAAGGGTTCTGGTTTGGCT 1161

Qy   359 IleThrArgLeuSerLeuCysAspLeuLysSerSerArgThrGlyValAlaPro 378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162 GCCACCTCCTGTACTCTCTTTAAAGTCACACAAGATGCTACACCGGGGTCTCGCTCTTGATAA 1221
```

```
Db  1222 ATAACAGCAAGTTTTTATGTGATTCTTAAATCAAGGCTTGATTCAAAGAAACTGGTTGTGTGGCACCA 1281
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   379 AspValPheAlaGlyLysAsnMetLysLeuLeuArgGluAspThrHisHisLeuValAsnTyrIle 398

Db  1282 GAGGTCTTCGCTGAAAACATGAAACCTCAGAGAGGACGACACCATCATTTTGGTGTCAACTATATT 1341
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrThrProTyrLeuValLysArgValAspGlu 418

Db  1342 CCCCAGGGTTCAAATGAGTTCAACTCTTTGAAGGAAACGTGGGTACTTAGGTGGTGGATGAA 1401
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspAspThrLeuArgLeuGlyGly 438

Db  1402 AAGCACAGAAGAAACTTTACCGTCGCGGCTCCACTTCCAAATGACTTTTGGATGAAGGA 1461
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisHisIleProSerProAlaLysLysVal 458

Db  1462 GTAGGACTTGTGCATTCAAACATAGCAACTGACCATATTCCAAGCCCTGCCAAGAAAGTA 1521
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIlleSerArgAsnGlyGluHis 478

Db  1522 CCAAGACTCCCCTGCCACAGACGACAAACCTGAAGCAGCTGTCATTACTAATGGGGAACAT 1581
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 10
US-10-236-417-183
; Sequence 183, Application US/10236417
; Publication No. US20040048256A1
; GENERAL INFORMATION:
; APPLICANT: Agee et al.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-442C
; CURRENT APPLICATION NUMBER: US/10/236,417
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: US60/318,120
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US60/318,430
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: US60/322,781
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/318,184
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US60/361,663
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US60/396,412
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: US60/322,636
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,817
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/322,816
; PRIOR FILING DATE: 2001-09-17
; PRIOR APPLICATION NUMBER: US60/323,519
; PRIOR FILING DATE: 2001-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Custom
; SEQ ID NO 183
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22)..(1582)
; US-10-236-417-183
DB:

Alignment Scores:
Pred. No.:
Score: 2480.00
Percent Similarity: 91.92%
Best Local Similarity: 91.92%
Query Match: 98.77%

Pred. No.:   1e-289   Length: 1650
Score: 2480.00   Matches: 478
Percent Similarity: 91.92%   Conservative: 0
Best Local Similarity: 91.92%   Mismatches: 0
Query Match: 98.77%   Indels: 42
                                    13   Gaps: 1

```
US-10-622-516-2 (1-478) x US-10-236-417-183 (1-1650)

Qy     1  MetProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProLysAspValGlyIleVal   20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22  ATGCCTGGATCACTTCCTTGAATGCAGGAGCTTGCTGGCCAAAAGATGTTGGAATTGTT   81

Qy    21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    82  GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT  141

Qy    41  GlyValAspAlaGlyLysTyrThrIleGlyLeuGlnAlaLysMetGlyPheCysThr      60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142  GGTGTAGATGCTGGGAAGTATACCATTGGCTTGCAGGCCAAAGATGGGCTTCTGCACA   201

Qy    61  AspArgGluAspIleAlaAsnSerLeuAsnMetLeuMetGluArgAsn              80
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   202  GATAGAGAAGATATTAACTCTCTTTGCAATGACTCTTATGGAGAGAAAT           261

Qy    81  AsnLeuSerTyrArgAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262  AACCTTTCCTATAGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAA  321

Qy   101  SerLysSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIle     120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAAGTGGGAATACAGATATA    381

Qy   121  GluGlyIleLeuAspThrThrArgAsnAlaCysTyrGlyThrAlaIleAlaValAlaVal  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382  GAAGGAATCGACACAACTAGGAATGCTTGTTATGGGACAGCAATCGCTGTGGCTGTT    441

Qy   141  AsnTrpIleGluGluSerSerTrpAsp-----------------------------     149
          |||||||||||||||||||||||||||
Db   442  AACTGGATTGAGGTCTCAGCTCTTGGGGATGGACGGGTATGCCCGTAGTTGCAGGAGATATT  501

Qy   149  ---------------------------------------                       149
          
Db   502  GCTGTATATGCCACAGGAAAATGCTAGACCTACACAGGTGTGGAGCAGTTGGAGCACTAGCTCTGCTA  561

Qy   150  ---------------GlyLeuArgGlyThrHisMetGlnHis                    158
                         ||||||||||||||||||||||||||
Db   562  ATTGGGCAAATGCTCTTTAATTTTGAACGAGGAGGGCTTCGTGGGACACATATCCAACAT   621

Qy   159  AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu  178
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   622  GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAACTC  681

Qy   179  SerIleGlyTyrLeuSerIleAspAspArgCysTyrSerValTyrCysLysLysIle     198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   682  TCCATACAGTGCTACCTGTCATTAGACGGCTGCTATTCTGTCTACTGCAAAAAGATC    741

Qy   199  HisAlaGlnIleTrpGlnLysGlyLysAsnAlaPheThrLeuAsnAspPheGlyPhe     218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   742  CATGCCCAGTGGCAGAAAGGGAAAAATGCATTTACCTTGAATGATTTTGGCTTC      801

Qy   219  MetIlePheHisSerProThrTyrCysLysLysSerLeuValAlaArgMetLeuLeu    238
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   802  ATGATCTTTCACTCACCATATTGTAAACTGTCTCAGCTGGATCGTTGCTG          861

Qy   239  AsnAspPheLeuAsnAsnGlnAsnAspArgLysAlaAsnSerIleTyrSerGlyLeuGluAla  258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862  AATGACTTCCTTAATGACAAGAATAGATAAAATAGTATCTAGTGGGCCCTGGAAGGCC   921

Qy   259  PheGlyAspValLysGluGluAspAlaAspThrTyrPheAsnAspArgAsnAlaPheMet  278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   922  TTTGGGGATGTTAAATTAGAAGAAGACGCAGATACTTTGATAGAGCGAATTGCATATG    981

Qy   279  LysAlaSerSerGluLeuPheSerGlnLeuValThrLysAspValSerAsnGln       298
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982  AAGGCTAGCTCTGAACTCTTCAGTCAGAAAACAAAGGACATCTTTACTGTATCAAATCAA 1041

Qy   299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnThrTyr  318
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1042  AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1101

Qy   319  SerProGlnLeuAlaGlyLeuLysArgIleGlyValPheSerTyrGlyLeuAla       338
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1102  TCACCTCAGCAAATTAGCACGGGAAGCAGAGAATTGGAGTGTTTCTTATGGTGGCT    1161

Qy   339  AlaThrLeuTyrSerLeuLeuValThrGlnAsnProAlaThrProGlySerAlaAlaLeuAspGlu  358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162  GCCACTCTGTACTCTCTTAAAGTCACAAACCCAGATACACCGGGGTCTGCTCTTGATAAA 1221

Qy   359  IleThrAlaSerLeuCysAspLeuLysSerArgArgThrGlyValAlaPro          378
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1222  ATAACAGCAAGTTTATGTGATCTTAAATCAAGAAGAACTGGTGTGGCACCA         1281

Qy   379  AspValPheAlaGluAsnMetLysSerArgGluAspThrHisHisLeuValAsnTyrIle  398
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1282  GATGTCTTCGCTGAAAACATGAAAGCTCAGAGAGACACCCATCATTTGGTCAACTATATT 1341

Qy   399  ProGlnGlySerIleAspSerLeuPheGluGlyTyrThrTrpTyrLeuValArgValAspGlu  418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1342  CCCCAGGGTTCAATAGATTCACTCTTTGAAGGATACACGTGGTACTTAGTTAGGGTGGATGAA 1401

Qy   419  LysHisLeuArgArgThrTyrAlaArgProThrProAsnAspAspThrLeuAspGlyGly  438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1402  AAGCACGAGAGAACTTACGCTCGGCCGCTCCCAAATGATGACACTTTGGATGGAGGA    1461

Qy   439  ValGlyLeuValHisSerAsnIleAlaThrGluHisSerIleProSerProAlaLeuLysLysVal  458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1462  GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521

Qy   459  ProArgLeuProAlaThrAlaAlaGluProGluLeuAlaAlaValIleSerAsnGlyGluHis  478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1522  CCAAGACTCCCTGCCACAGCAGCAGAAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT 1581
```

US-10-236-417-185

Alignment Scores:

Pred. No.:

Score:	2480.00
Percent Similarity:	91.92%
Best Local Similarity:	91.92%
Query Match:	98.77%

1e-289

Length:	1650
Matches:	478
Conservative:	0
Mismatches:	0
Indels:	42
Gaps:	1

DB:

US-10-622-516-2 (1-478) x US-10-236-417-185 (1-1650)

```
Qy    1   MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1   ATGCCTGGATCACTTCCTTTGAATCAGAGAAGTGTTGGCCCAAAAGATGTGGGAATTGTT   81

Qy   21   AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlyLysTyrAsp   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22   GCCCTTGAGACTCTATTTCTCCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT   141

Qy   41   GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyLnAlaLeuMetGlyPheCysThr   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82   GGCCCTTGAGACTCTATTTCTCCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT   201

Qy   61   AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142   GGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGCCAGGCCCAAGATGGGCCTTCTGCACA   261

Qy   81   AsnLeuSerTyrAspCysIleGlyArgLeuGlyValGlyThrGluThrIleIleLeuAspLys  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202   GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT   321

Qy  101   SerGlyLysSerValLysThrLeuMetGlyLeuPheGluGlyLysAsnThrArgIle  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262   AACCCTTTCCTATGATTTCCATTGGGCGCCGGCTGGGAAGTTGGAACAGAGACAAATCATCATCGACAA   381

Qy  121   GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322   TCAAAAGTCTGTGAAGACTAATTTGATGCAGACTGTCTGGGAAGTGCTCTGGGAATACAGATATA   441

Qy  141   AsnTrpIleGluSerSerSerTrpAsp---------------------------------  149
          |||||||||||||||||
Db  382   GAAGGAAATCGACAACTAATGCATGCTATCATGGAGGCACAGCTCGCTGTCTTCAAATGCTGTT   501

Qy  149   ------------------------------------GlyLeuArgGlyThrHisMetGlnHis  158
                                              ||||||||||||||||||||||||||
Db  442   AACTGGATTGAGTCCCAGCTCTTGGGATGGACGGTARTGCCCTGGTAGTTGCAGGAGATATT   561

Qy  150   -----------------------------------GlyLeuArgGlyThrHisMetGlnHis  158
          
Db  502   GCTGTATATGCCACAGGAAATGCTAGACCTTACAGGTGGTGGAGTTGGAGCAGTAGCTCTGCTA   621

Qy  159   AlaTyrAspPheTyrLysProGluLeuValAspGlnAlaGluLeu  178
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  562   ATTGGGGCCAAATGCTCCTTTAATTTTGAACGAGGGCCTTCGTGGGACACATATGCAACAT   681

Qy  179   SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysGluIle  198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622   GCCTATGATTTTTACAAGCCTGAGTTAGTAGATCAGGCTGAACTC   741

Qy  199   HisAlaGlnTrpGlnLysGlyAsnAs.pAspPheThrLeuAsnAspPheGlyPhe  218
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  682   TCCATACAGTGCTACCTCAGTGCATTAGACCGCTGCTATTCTGTTTCTACTGCAAAAGAATC   801

Qy  219   MetIlePheHisSerProTyrCysLysLysGluLeuIleAlaArgMetLeuLeu  238
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  742   CATGCCCAGTGGCAGAAAGGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCCTTC   861

Qy  239   AsnAspPheLeuAsnAspGlnAsnArgAspAspPheAsnSerIleTyrSerGlyLeuGluAla  258
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  802   ATGATCTTTCACTCACCATCATATTGTAAACTGGGTTCAGAAAATCTCTAGCTCGGAGTTGCTG   921

Qy  259   PheGlyAspValLysLeuGluGluAspThrTyrPheAspArgAspValGluLysAlaPheMet  278
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  862   AATGACTTCTTAAATGACCAGAATAGAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC   981
```

```
Db  922   TTTGGGGATGTTAAATTAGAAGAACACCTTACTTTGATAGATGGTTTCTCTTGTGGAGAAGGCATTTATG   981

Qy  279   LysAlaSerSerGluLeuPheSerGlnTyrGlyLysThrLysSerAlaSerLeuLeuValSerArgAsnGln  298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982   AAGGCTAGCTCTGAACTCTTCAGTCAGTATGGAAAAACAAAAGGCATCTTTACTTGTATCAAAAATCAA   1041

Qy  299   AsnGlyLysAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuLeuAlaGlnTyr  318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042   AATGGAAAATATGTACACATCTTCAGTTATGGTTCCCTTGGATCGTGTTCTAGCACAGTAC   1101

Qy  319   SerProGlnGlnLeuAlaGlyLysValArgIleGlyLysValPheSerTyrGlySerGlyLeuAla  338
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102   TCACCTCAGCAATTAGCAGGGAAGGAAGAAATTGGGAGTGGTTTTCTTAGGGTTCGGGTTTGGCT   1161

Qy  339   AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuLeuAspGlu  358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162   GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCCGGGTCTGCCTCTTGATAA   1221

Qy  359   IleThrAlaSerLeuCysAspGlnLeuGlySerArgLeuAspSerArgThrGlyValAlaPro  378
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222   ATAACAGCAAGTTTAGTCGATCTTAAATCAAGGCCTTGAATTCAAGAACTGGTGGTGGCACCA   1281

Qy  379   AspValPheAlaGluAsnMetLysLeuArgGluArgGluAspProThrHisHisIleLeuValAsnTyrIle  398
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282   GATGTTCTTCGCTGAAAACATGAAGCTCAGAAGAGAACACCCATCATTCATTTGGTCAACTATATT   1341

Qy  399   ProGlnGlySerIleAspSerSerLeuPheGluGlyTyrThrArgTyrLeuValAlaArgValAspGlu  418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1342   CCCCAGGGTTCAATAGATTCACTCTTTGAAGGATACACGCGGTACTTAGTTGCGGTGGATGAA   1401

Qy  419   LysHisIleArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGly  438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1402   AAGCACAGAAGAACTTACGCTCGGCGGCCTCCAAATGATGACACTTTGGATGGAAGGA   1461

Qy  439   ValGlyLeuValHisSerAsnIleIleAlaThrGluHisIleLeuProSerProAlaLysGlyLysVal  458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1462   GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCCTGCCAAGAAAGTA   1521

Qy  459   ProArgLeuProAlaThrAlaAlaLeuProGluAlaAlaValIleSerSerAsnGlyGluHis  478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1522   CCCAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTAGTAATGGGGAACAT   1581
```

RESULT 12

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom

;  SEQ ID NO 189
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (22)..(1582)

US-10-236-417-189
;  LOCATION: (22)..(1582)

US-10-622-516-2 (1-478) x US-10-236-417-189 (1-1650)

Alignment Scores:
Pred. No.:
Score:                  2480.00          Length:        1650
Percent Similarity:     91.92%           Matches:        478
Best Local Similarity:  91.92%           Conservative:     0
Query Match:            98.77%           Mismatches:       0
DB:                     13               Indels:          42
                                         Gaps:             1
```

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal    20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 ATGCCTGGGATCACTTCCTTTGAATGCAGAAGCTTGCCCAAAAGATGTTGGGATTGTT     81

Qy   21 AlaLeuGluIleTyrPheProSerGlnThrTyrValValAspGlnAlaGluLeuGluTyrAsp    40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82 GCCCTTGAGATCTATTTTCCTTCCCAATAGTTGTGGATCAGAGTTGGAAAAATATGAT    141

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyTyrLeuGlyLysAlaLysMetGlyPheCysThr    60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142 GGTGTAGATGCTGGGAAGTATACCATTGGCTTGGGGTAAAGATGGCTTCTGCACA     201

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn    80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT    261

Qy   81 AsnLeuSerThrArgAspCysIleGlyArgLeuLeuValGlyThrGluThrIleIleAspLys    100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262 AACCTTTCCTACTAGATTGCATTGGGCGGCTCCTGGAAGTTGGAACAGAGACAATCATCGACAAA    321

Qy  101 SerLysSerValLysThrArgAspCysTyrGlyLeuLeuValGlyGlyLeuTyrAsnThrAspIle    120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322 TCAAAGTCTGTGAAGACTAGATTTGATGGACTGTTTGAGATACAGATATA     381

Qy  121 GluGlyIleIleAspThrThrAsnAlaCysTyrGlyLeuThrAlaAlaValAlaValAsp    140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382 GAAGGAAATCGACACAACTAATGCATGCTATGGAGGCACAGCTGCTGTTGTT     441

Qy  141 AsnTrpIleGluSerSerSerTrpAsp---                                149
          |||||||||||||||||||||||||
Db  442 AACTGGATTGAGTCCAGCTCTTGGGATGGGACGGGTATGCCCTGGTTAGTTGCAGGAGATATT    501

Qy  149 ---------                                                     149
Db  502 GCTGTATATGCCACAGGAAATGCTAGACCTACAGGTGGAGCAGTAGCTCTGCTA     561

Qy  150 ---GlyLeuArgGlyThrHisMetGlnHis                                158
                 |||||||||||||||||||
Db  562 ATTGGGGCCAAATGCTCCTTTAATTTGAACGAGGGCTTCGTGGGACACATATGCAACAT    621

Qy  159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu    178
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGAAAACTC    681

Qy  179 SerIleGluGlnCysTyrLeuSerArgAlaLeuAspArgCysTyrSerValTyrCysLysLysIle    198
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682 TCCATACAGTGCTACCTCAGTCGATTAGACCGCTGCTATTCTGTCTACTGCAAAAAGATC    741

Qy  199 HisAlaGlnThrTrpGlnLysGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe    218
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742 CATGCCCAGTGGCAGAAAGGGAAATGATAAAGATTTTACCTTGAATGAATTTGGCTTC    801
```

```
RESULT 13
US-10-236-417-193
Sequence 193, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
;  APPLICANT:
;  APPLICANT: Agee et al.
;  TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;  FILE REFERENCE: 21402-442C
;  CURRENT APPLICATION NUMBER:
;  CURRENT FILING DATE:       2003-01-06
;  PRIOR APPLICATION NUMBER:  US/10/236,417
;  PRIOR FILING DATE:         2001-09-01
;  PRIOR APPLICATION NUMBER:  US60/318,120
;  PRIOR FILING DATE:         2001-09-10
;  PRIOR APPLICATION NUMBER:  US60/318,430
;  PRIOR FILING DATE:         2001-09-10
;  PRIOR APPLICATION NUMBER:  US60/322,781
;  PRIOR FILING DATE:         2001-09-17
;  PRIOR APPLICATION NUMBER:  US60/318,184
;  PRIOR FILING DATE:         2001-09-07
;  PRIOR APPLICATION NUMBER:  US60/361,663
;  PRIOR FILING DATE:         2002-03-05
```

```
Qy  219 MetIlePheHisSerProTyrCysLysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu    238
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802 ATGATCTTTCACTCACCATATTGTAAAACGTTAAAATCTCTAGCTCGGATGTTGCTG    861

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla    258
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATACTCTATAGTGGCTTGGAAGCC    921

Qy  259 PheGlyAspValLysLeuGluAspThrThrPheAspArgAspValGlyLysAlaPheMet    278
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922 TTTGGGGATGTTAAAATTAGAAGACACCTACTTGATAGAGAGGTGGGAAAGGCATTTATG    981

Qy  279 LysAlaSerGluGluLeuPheSerGlnLeuLysThrLysAspSerLeuLeuValSerSerAsnGln    298
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982 AAGGCTAGCTCTGAACTCTTCAGTCAGAAACAAGGCATCTTTACTTGTTATCAAATCAA    1041

Qy  299 AsnGlyAsnMetTyrThrSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr    318
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042 AATGGAAATATGTACACATCTTCAGTATATGGTTCCCTTGCATCTGTTCTAGCACAGTAC    1101

Qy  319 SerProGlnGlnLeuValAlaGlyIleGlyValPheSerThrTyrGlySerGlyLeuAla    338
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102 TCACCTCAGCAATTACCAGGGAAGGAAGAATTGGAGTGTTTTCTTACGTATGGTTCAGGGCT    1161

Qy  339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys    358
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162 GCCACTCTGTACTCTCTTAAAGTCACACAGGATGCTACACCGGGGTCTGCTCTTGATAAA    1221

Qy  359 IleThrAlaSerLeuCysAspAlaLysLysSerArgLeuAspSerArgThrGlyValAlaPro    378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222 ATAACAGCAAGTTTATGTGATGCTAAAAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA    1281

Qy  379 AspValPheAlaGlyLeuAsnMetLysLeuGlnGlyAspThrHisHisLeuLeuAsnTyrIle    398
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282 GATGTCTTCGCTGGACTAAACATGAAGCTTCAGGGAGACACCCATCATTTGCTAAACTATATT    1341

Qy  399 ProGlyGlnGlySerIleAspSerLeuPheGluGlyLeuThrThrProTyrLeuValArgValAspGlu    418
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1342 CCCGGACAGGGTTCAATAGATTCACTCTTTGAAGGACTAACGACTCCTTAGTTAGGGTTGATGAA    1401

Qy  419 LysHisIleArgArgThrTyrAlaArgProThrProAsnAspAspHisLeuAspGluGly    438
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1402 AAGCACAGAAGAACTTACGCTCGGCCCACTCCAAATGACGACCATTTGGATGAAGGA    1461

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleIleProSerProAlaLysLysLysVal    458
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1462 GTAGGACTTGTGCATTCAAACATAGCAACTGAACATATTATCCCAAGCCCTGCCAAGAAAGTA    1521

Qy  459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleIleSerAsnGlyGluHis    478
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1522 CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGCTGTCATTATAAGTAATGGGGAACAT    1581
```

```
;   PRIOR APPLICATION NUMBER: US60/396,412
;   PRIOR FILING DATE: 2002-07-17
;   PRIOR APPLICATION NUMBER: US60/322,636
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,817
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/322,816
;   PRIOR FILING DATE: 2001-09-17
;   PRIOR APPLICATION NUMBER: US60/323,519
;   PRIOR FILING DATE: 2001-09-19
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 341
;   SOFTWARE: Custom
;   SEQ ID NO 193
;   LENGTH: 1650
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22)..(1582)
US-10-236-417-193
DB:

Alignment Scores:
Pred. No.:              1e-289
Score:                  2480.00
Percent Similarity:     91.92%
Best Local Similarity:  91.92%
Query Match:            98.77%
DB:                     13
```

```
US-10-622-516-2 (1-478) x US-10-236-417-193 (1-1650)

                        Length:        1650
                        Matches:       478
                        Conservative:  0
                        Mismatches:    0
                        Indels:        42
                        Gaps:          1
```

```
Qy    1  MetProGlySerLeuProLeuAsnAlaGlyValAlaCysTrpProLysAspValGlyIleVal  20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22  ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGCCCAAAAGAATGTTGGGGATTGTT  81

Qy   21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp  40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82  GCCCTTGAGATCTATTTCCCTTCTCAAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT  141

Qy   41  GlyValAspAlaGlyLysTyrThrIleGlyLysGlnAlaLysMetGlyPheCysThr  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  142  GGTGTAGATGCTGGGAAGTATACCATTGGTTGGGCCAAGATGGGCTTCTGCACA  201

Qy   61  AspArgGluAspIleAsnAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn  80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  202  GATACAGAAGAGATAATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT  261

Qy   81  AsnLeuSerTyrArgAspCysIleGlyArgLeuGluValGlyThrGlyThrIleIleAspPro  100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  262  AACCTTTCCTATAGATTGCATTGGGCGGCTGGAAGTTGGAACAGGAACAATCATCGACAAA  321

Qy  101  SerLysSerValLysThrAspThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  322  TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA  381

Qy  121  GluGlyIleAspThrThrAsnAlaCysTyrGlyIleGlyThrAlaLeuAlaValPheAsnAlaVal  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  382  GAAGGAAATCGACACAACTAATGCATATGGGAGGCACAGCTGCTGTCTTCTTCAAAGCTGTT  441

Qy  141  AsnTrpIleGluSerSerSerTrpAsp----------------------  149
         |||||||||||||||||||||||
Db  442  AACTGGATTGAGTCCAGCTCTTGGGATGGACCGGTATGCCCTGGTAGTTGCAGGAGATATT  501

Qy  149  ------------------------------------------------  149
Db  502  GCTGATATGCCACAGGAAATGCTAGACCTACAGGGTGGAGTTGGAGCAGTAGCTCTGCTA  561

Qy  150  ------------------GlyLeuArgGlyThrHisMetGlnHis  158
                           ||||||||||||||||||
Db  562  ATTGGGCCAAATGCTCCTTTAATTTTGAACAGGAGGGCTTCCTGGGACACATATGCAACAT  621

Qy  159  AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu  178
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  622  GCCTATGATTTTTACAAGCCTGATATGCTTAGTGAATATCCTATAGTAGATGGGAAACTC  681

Qy  179  SerIleGlnCysTyrLeuSerArgAlaLeuAspArgCysTyrSerValTyrCysGlyLysIle  198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  682  TCCATACAGTGCTACCTCAGTCGAGCATTAGACCGCTGCTATTCTGTCTACTGTGGAAAGATC  741

Qy  199  HisAlaGlnThrGlnLysGlyLysAsnAspLysAspPheThrLeuAsnAspPheGlyPhe  218
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  742  CATGCCCAGTGCCAGAAAGGAAAGAATGATAAAGGATTTTTACCTTTGAATGATTTTGGCTTC  801

Qy  219  MetIlePheHisSerProTyrCysGlyLysValGlnLysSerLeuAlaArgMetLeu  238
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  802  ATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCCTCAGCTCGGAATGTTG  861

Qy  239  AsnAspPheLeuAsnAspGlnAsnArgAspProLysAsnSerIleTyrSerGlyValGluGluAla  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  862  AATGACTTCCTTAATGACCAGAATAGAGATCCAAAAAATAGTATCTATAGTGGAGTAGAAGCC  921

Qy  259  PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysAlaPheMet  278
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  922  TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGTTGGGAAGGCATTTATG  981

Qy  279  LysAlaSerSerGluLeuPheSerGlnLysThrIleGlyValPheSerGlySerGlyLeuAla  298
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  982  AAGGCTAGCTCTGAACTCTTCAGTCAAAAGACAATTGGAGTGTTTTCTGGTTCTGGCT  1041

Qy  299  AsnGlyAsnMetTyrThrSerValTyrGlySerLeuAlaSerValLeuLeuAlaGlnTyr  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1042  AATGGAAATATGTACACATCTTCAGTATAGGTTGCCATCCTGTTCTAGCACAGTAC  1101

Qy  319  SerProGlnGluLeuAlaGlyLysValArgIleGlyValPheSerArgTyrGlySerGlyLeuAla  338
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1102  TCACCTCAGCAATTAGCAGGGAAAGAGAATTGGAGTGTTTTCTTAAGGTATGGTTCTGGGCT  1161

Qy  339  AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuLeuAspLys  358
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1162  GCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGGTCTGCTCTTCTTGATAAA  1221

Qy  359  IleThrArgLeuSerLeuCysAspSerLeuLysSerSerArgLeuAspSerArgThrGlyValAlaPro  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1222  ATAACAGCAAGTTATTGACTTTAAATCAAGGCTTGATTCAAGAACTGGTGTGGCACCA  1281

Qy  379  AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuLeuValAsnTyrIle  398
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1282  GATGTTTTCGCTGAAAACATGAAGCTTCGAGAAGACACCCATCATTTGCTGGTCAACTATATT  1341

Qy  399  ProGlnGlySerIleAspSerLeuPheGluGlyLeuTyrThrTrpTyrLeuValValArgValAspGlu  418
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1342  CCCCAGGGTTCAATAGATTCACTCTTTGAAGGACTGTGGTACTTAGTTGAGTTGGATGAA  1401

Qy  419  LysHisArgArgThrTyrAlaAsnArgArgProThrProAsnAspAspThrLeuLysLysGly  438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1402  AAGCACAGAAGAACTTACGCTCCGGCGTCCCACTCCAAATGATGACACCTTGAAGAAGGA  1461

Qy  439  ValGlyLeuValHisSerAsnIleAlaThrGluHisSerIleProSerProAlaLysLysVal  458
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1462  GTAGGACTTGTGCATTCAAACATTAGCAACAGCATATTCCAAAGCCCTGCCAGAAAGTA  1521

Qy  459  ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis  478
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1522  CCAAGACTCCCTGCCACAGCAGCAGAACCTGAAGCAGTCGTTCATTAGTAATGGGGAACAT  1581
```

```
RESULT 14
US-10-236-417-197
;   Sequence 197, Application US/10236417
;   Publication No. US20040048256A1
;   GENERAL INFORMATION:
;   APPLICANT: Agee et al.
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-442C
;   CURRENT APPLICATION NUMBER: US/10/236,417
;   CURRENT FILING DATE: 2003-01-06
```

```
;  PRIOR APPLICATION NUMBER: US60/318,120
;  PRIOR FILING DATE: 2001-09-01
;  PRIOR APPLICATION NUMBER: US60/318,430
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: US60/322,781
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/318,184
;  PRIOR FILING DATE: 2001-09-07
;  PRIOR APPLICATION NUMBER: US60/361,663
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US60/396,412
;  PRIOR FILING DATE: 2002-07-17
;  PRIOR APPLICATION NUMBER: US60/322,636
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,817
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/322,816
;  PRIOR FILING DATE: 2001-09-17
;  PRIOR APPLICATION NUMBER: US60/323,519
;  PRIOR FILING DATE: 2001-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 341
;  SOFTWARE: Custom
;  SEQ ID NO 197
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (22)..(1582)
US-10-236-417-197

Alignment Scores:
Pred. No.:
Score:                     1e-289
Percent Similarity:        2480.00
Best Local Similarity:     91.92%
Query Match:               91.92%
DB:                        98.77%
                           13

US-10-622-516-2 (1-478) x US-10-236-417-197 (1-1650)

                           Length:        1650
                           Matches:       478
                           Conservative:  0
                           Mismatches:    0
                           Indels:        42
                           Gaps:          1
```

```
Qy     1 MetProGlySerLeuProLeuAsnAlaGlyAlaCysTrpProGlyLysAspValGlyIleVal 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22 ATGCCTGGATCACTTCCTTTGAATGCAGGAGCTTGCTGCCCAAAAGATGTTGGGGATTGTT 81

Qy    21 AlaLeuGluIleTyrPheProSerGlnTyrValAlaSpGlnAlaGluGluGlyLysTyrAsp 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    82 GCCCTTGAGATCTATTTTCCTTCTCAATATGTGACAGATGCAGAGTTGGAAAAATATGAT 141

Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysGlyThr 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 GGTGTAGATGCTGGGAAGTATACCATTGGCCTTGGGCAGGCCAAGATGGGCTTCTGCACA 201

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 261

Qy    81 AsnLeuSerTyrAspCysIleGlyTyrArgLeuGluValGlyIleTyrGluThrIleIleAspLys 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 AACCTTTCCTATGATTGCATTGGGTACCGGCTGGAAGTTGGAATCTACGAGACAATCATCGACAAA 321

Qy   101 SerLysSerValLysThrLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 TCAAAGTCTGTGAAGACTAAATTGATGCAGCTTTTTGAAGAGTCTGGGAATACAGATATA 381

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyLysGlyThrAlaAlaValPheAsnAlaVal 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382 GGAGGAATCGACACAACTAATGCATGCTATGGGAAGGGCACAGCTGCTGTTTTCAATGCTGTT 441

Qy   141 AsnTrpIleGluSerSerSerTrpAsp--------------------- 149
           |||||||||||||||||||||||||||
Db   442 AACTGGATTGAGTCCAGCTCTTGGGATGGGACGTATGCCCTGCTGTAGTTGCAGAGATATT 501
```

```
Qy   149 --------------------------------------------------------- 149
Db   502 GCTGTATATGCCACAGGAAATGCTAGAACCTACACGGTGGGAGTTGGAGCAGTCTGCTA 561

Qy   150 ----------------------------------GlyLeuArgGlyThrHisMetGlnHis 158
                                              |||||||||||||||||
Db   562 ATTGGGGCCAAATGCTCCTTTAATTTTGAACAGAGGCTTCGTGGACACTATGGCAACAT 621

Qy   159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleLeuValAspGlyLeu 178
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   622 GCCTATGATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGACTC 681

Qy   179 SerIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   682 TCCATACAGTGCTACCTGTCAGCATTAGACCGCTGTTATTCTGTCTACTGCAAAAAGATC 741

Qy   199 HisAlaGlnIleTrpGlnLysSGluGlyAsnAspLysAspPhePheLeuAsnAspPhePheGlyPhe 218
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   742 CATGCCCAGTTGCAGAAGAGCGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 801

Qy   219 MetIlePheHisIleSerProTyrCysLysLysValGlnAsnLeuLysSerLeuAsnLeu 238
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   802 ATGATCTTTCACATCAGCATTCGTAAATGCTGGTTCAGGAAAATCTCTAGCTCGGATTGTG 861

Qy   239 AsnAspPheLeuAsnAsnGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla 258
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   862 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGCCTGGAAGCC 921

Qy   259 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuLeuAlaGlnTyr 278
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   922 AATGGAAATATGTACACATCTTCAGTATATGGGTCCCTGGCATCGTGTTCTAGCAGTAC 981

Qy   279 PheGlyAspValLysGluGlyAspThrTyrPheAspArgAspValGlyLysGlyAlaPheMet 298
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   982 TTTGGGGATGTTAAATTAGAAGGAGACACCTACTTTGATAGAGATGTGGGAAGGGCATTTATG 1041

Qy   299 LysAlaSerSerGluLeuPheSerGlnLysThrSerAlaSerLeuArgGlyLeuGluAla 318
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1042 AAGGCTAGCTCTGAACTCTTTCAGAAAACAAGCGCATCTTTAAGGTTTTGGTTGGGGCT 1101

Qy   319 SerProGlnGlnLeuAlaGlyLysAspArgIleGlyValPheSerGlyLeuGlyLeuAla 338
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1102 TCACCTTCAGCAAATTAGCAGGGAAGAGAGAATTGGAGTGTTTTCTAAGGTTTCTGGGTTGGCT 1161

Qy   339 AlaThrLeuTyrSerLeuValThrGlnAsnAspIleAsnProGlySerAlaLeuAspTyr 358
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1162 GCCACTCTGTACTCTCTTGTAACTCAAAATGATATTAACCCGGGGTCTGCTTGATAAA 1221

Qy   359 IleThrArgAlaSerLeuCysAspLeuIleLysSerArgAlaSerArgThrGlyValAlaPro 378
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1222 ATAACAAGAGCAAGTTTATGTGATCTTATAAAATCAAGGCGTTGATTCAAGAAACTGGTGTGGCACCA 1281

Qy   379 AspValPheAlaGluAspAsnMetLysLeuValArgGluAspThrHisHisLeuValAsnTyrIle 398
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1282 GATGTTTCGCTGAAAGATAACATGAAGCTTGCAGAGGACCACCCATCATTTTGTCAACTATATT 1341

Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyTyrTrpTyrLeuValArgValAspGlu 418
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1342 CCCCAGGGGTTCAATGATTCACTCTTTGAAGGAACGTGGTACTTGGTGAGGGTGGATGAA 1401

Qy   419 LysHisArgGlyArgThrTyrAlaArgArgProThrProAsnAspAspThrProLeuAspGlyGly 438
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1402 AAGCACAGAGGACGTACCTACGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGGTGAAGGA 1461

Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal 458
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1462 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA 1521

Qy   459 ProArgLeuProAlaThrAlaAlaAlaAlaValIleSerAsnGlyGluHis 478
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1522 CCAAGACTCCCTGCCACAGCAGCAGCAGCTGTCATTAGTAATGGGGAACAT 1581
```

```
RESULT 15
US-10-236-417-201

Sequence 201, Application US/10236417
Publication No. US20040048256A1
GENERAL INFORMATION:
APPLICANT: Agee et al.
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-442C
CURRENT APPLICATION NUMBER: US/10/236,417
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: US60/318,120
PRIOR FILING DATE: 2001-09-01
PRIOR APPLICATION NUMBER: US60/318,430
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: US60/322,781
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/318,184
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US60/361,663
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US60/396,412
PRIOR FILING DATE: 2002-07-17
PRIOR APPLICATION NUMBER: US60/322,636
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,817
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/322,816
PRIOR FILING DATE: 2001-09-17
PRIOR APPLICATION NUMBER: US60/323,519
PRIOR FILING DATE: 2001-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 341
SOFTWARE: Custom
SEQ ID NO 201
  LENGTH: 1650
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (22)..(1582)

US-10-236-417-201

Alignment Scores:
Pred. No.:            1e-289
Score:               2480.00
Percent Similarity:   91.92%
Percent Identity:     91.92%
Best Local Similarity: 98.77%
Query Match:          13

DB: US-10-236-417-201

Length:       1650
Matches:       478
Conservative:    0
Mismatches:      0
Indels:         42
Gaps:            1
```

```
US-10-622-516-2 (1-478) x US-10-236-417-201 (1-1650)

Qy      1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 TCAAAGTCTCTGAGAGACTAATTTGATGCAGCTGTGTTTGAAGAGTCTGGGAATACAGATATA 381

Qy     21 AlaLeuGlyIleTyrPheProSerGlnThrTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 GAAGGAATCGGACACAACTAATGCACTATTTTGAGGCACAGCAGGCAGCTCGTCTTTCAAATGCTGTT 441

Qy     41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyAlaLysMetGlyPheCysThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 AACTGGATTGAGTCCAGCCTCTTGGGATGGACGGTATGCCCTGGTAGTTGCAGGAGAGATATT 501

Qy     61 AspArgGluAspIleAsnSerLeuCysMetThrValGlnAsnLeuMetGluArgAsn 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 GCTGTATATGCCACACGAAAATGCTAGACCTACAGGTGGAGTTGGAGCAGTAGCTCGTCTA 561

Qy     81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyTyrThrGluThrIleIleAspGlu 100
          ||||||||||||||||||||                ||||||||||||||||||||
Db    562 ATTGGGGCCAAAATGCTCCTTTAATTTTTGAAGCAGAGGGTTCGTGGGACACATATGCACAT 621

Qy    101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle 120
          |||||||||||||||||||||||              |||||||||||||||||||||
Db    622 GCCTATGATTTTTACAAAGCCCGATATGCTATCCTATAGTAGATGGGAAACTC 681

Qy    121 GluGlyIleLeuAspThrArgAsnAlaCysTyrGlyLeyThrAlaLeuAlaValPheAsnAlaVal 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682 TCCATACAGTGCTTACCTCAGTGCATTAGACCCGTGCTATTCGTGTCTACTGCAAAAAGATC 741

Qy    141 AsnTrpIleGluSerSerSerTrpAsp 149
          |||||||||||||||||
Db    742 CATGCCCAGTGCGCAGAAAGAGGGAAATTTTACCTTGAAGAATGATTTTGGCTTC 801

Qy    149 ---------GlyLeuArgGlyTyrThrHisMetGlnHis 158
                   |||||||||||||||||||
Db    802 ATGGATCTTTCACTCACCATTATTGTAAACTGGTTCAGAAATCTCTAGCTCTGGGATTGTGCTG 861

Qy    159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
          |||||||||||||||||||                 |||||||||||||||||||
Db    862 AATGACTTCCTTAATGACCAGAATAGGAGATAAAAAGTATATCTATAGTGGCCCTGGAGCC 921

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAspAlaArgCysTyrTyrSerValTyrCysLysLysIle 198
          ||||||||||||||||||||||                |||||||||||||||||
Db    922 TTTGGGGATGTTAAATTAGAAGACACCTACTTTGATAGAGATGGGAGAAGGCCATTTATG 981

Qy    199 HisAlaGlnInTrpGlnLysGlyLysAsnAspLysAspPheTheLeuAsnAspPheGlyPhe 218
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    982 AAGGCTAGCTCTGGAACTCTTTCAGTCACAGAACAAAGGCATCTTTACTTGTATCAAAATCAA 1041

Qy    219 MetIlePheHisHisSerProTyrCysLysLysLysValGlnLysSerLeuAlaArgMetLeuLeu 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1042 AATGGAAAATATGTACACATCTTCAGATATATGGGTTCCCTTGCATCTGTTCTAGCACAGTAC 1101

Qy    239 AsnAspPheLeuAsnAspGlnAsnArgArgLysAsnSerIleTyrSerGlyLeuGluAla 258
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1102 TCACCTCACGCAAATTAGCAGGAGGAGAGAATTGGAGTGTTTTCTTAAGGGTTCTGGTTTGGCT 1161

Qy    259 PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGlyLysLysAlaPheMet 278
          ||||||||||||||||||||                |||||||||||||||||||
Db   1162 GCCACTCTGTACTCTCTTAAAGTCACACAAGATGTACACCGGGTCTCGCTCTTGATAA 1221

Qy    279 LysAlaSerSerGluLeuPheSerGlnLysThrLysLysAlaSerIleLeuValSerAsnGln 298
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 ATAACAGCAAGTTTAATGTCGATCTTAAATCAAGGGCTTGATTCCAAGAAACTGGGTGTGCCACCA 1281

Qy    299 AsnGlyAsnMetTyrThrSerSerValTyrCysSerLeuAlaSerValAlaLeuGlnTyr 318
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 GATGTCTTCCCTGAAAAACATGAAAGCTCAGAGAGGACACCCATCAATTTGGTCAACTATATT 1341

Qy    319 SerProGlnGlnLeuAlaGlyLysAsnArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1342 CCCCAGGGTTCAATAGATTCACTCCTTTCGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1401

Qy    339 AlaThrLeuTyrSerLeuValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
          ||||||||||||||||||||||||||||||||||||||||||||
Db   1402 ATGACTTCTGTACTCTCTTAAAGTCACAAGATGTACACCGGGTCTCGCTCTTGATAA 1221

Qy    359 IleThrAlaSerLeuCysAspLeuLysSerSerArgLeuAspSerArgThrGlyValAlaPro 378
          |||||||||||||||||||||||||||||||||||||||||||
Db   1222 ATAACAGCAAGTTTAATGTCGATCTTAAATCAAGGGCTTGATTCCAAGAAACTGGGTGTGCCACCA 1281

Qy    379 AspValPheAlaGlyAsnMetLysLeuArgGluAspThrHisHisGluLeuValAsnTyrIle 398
          |||||||||||||||||||||||||||||||||||||||||||
Db   1282 GATGTCTTCCCTGAAAAACATGAAAGCTCAGAGAGGACACCCATCAATTTGGTCAACTATATT 1341

Qy    399 ProGlnGlySerIleAspSerLeuLeuSerLysGluArgGlyThrTrpTyrLeuValArgValAspGlu 418
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   1342 CCCCAGGGTTCAATAGATTCACTCCTTTCGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1401

Qy    419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGluGly 438
          ||||||||||||||
Db   1402 CCCCAGGGTTCAATAGATTCACTCCTTTCGAAGGAACGTGGTACTTAGTTAGGGTGGATGAA 1401
```

```
Db   1402 AAGCACACAGAAGAACTTACGGCTCGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGA 1461

Qy    439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysLysVal   458
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1462 GTAGGACTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAGTA   1521

Qy    459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis   478
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1522 CCAAGACTCCCTGCCACAGCAGCAAGAACCTGAGCAGCTGTCATTAGTAAATGGGGAACAT 1581
```

Search completed: June 24, 2004, 17:13:40
Job time : 577 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:29:18 ; Search time 115 Seconds
          (without alignments)
          2306.665 Million cell updates/sec

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV..........PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0

Searched: 682709 seqs, 277475446 residues

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10622516/runat_23062004_162634_673/app_query.fasta_1.647
-DB=issued_Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10622516 @CGN_1_1.105 @runat_23062004_162634_673 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Issued_Patents NA:*
1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

Database :

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2511	100.0	2002	4	US-09-819-993-1	Sequence 1, Appli
2	2511	100.0	2002	4	US-10-193-295-1	Sequence 1, Appli
3	2089	83.2	1824	1	US-08-305-505-1	Sequence 1, Appli
4	839.5	33.4	28001	4	US-09-819-993-3	Sequence 3, Appli
5	839.5	33.4	28001	4	US-10-193-295-3	Sequence 3, Appli
6	771	30.7	615	4	US-09-370-838-245	Sequence 245, App
7	703	28.0	500	4	US-09-370-838-128	Sequence 128, App
8	686	27.3	506	4	US-09-401-064-187	Sequence 187, Appl
9	585	23.3	4775	3	US-09-306-595C-1	Sequence 1, Appli
10	585	23.3	4775	4	US-09-925-388-1	Sequence 1, Appli
11	530.5	21.1	635	4	US-08-998-416-730	Sequence 730, Appl
12	395	15.7	472	4	US-09-833-381-1648	Sequence 1648, Ap
13	315	12.5	1167	4	US-09-107-532A-1715	Sequence 1715, Ap
14	301.5	12.0	10813	4	US-08-956-171E-364	Sequence 364, App
15	284.5	11.3	1311	4	US-09-134-001C-1458	Sequence 1458, Ap
16	282	11.2	1350	4	US-09-134-000C-2039	Sequence 2039, Ap
17	268	10.7	305	4	US-09-313-294A-4313	Sequence 4313, Ap
18	262.5	10.5	448	4	US-09-833-381-1645	Sequence 1645, Ap
19	255.5	10.2	288	5	US-09-313-294A-5701	Sequence 5701, Ap
20	244	9.7	18436	4	US-09-961-527-87	Sequence 87, Appl
21	242	9.6	307	4	US-09-833-381-744	Sequence 744, App
22	209.5	8.3	951	4	US-09-241-750-1	Sequence 1, Appli
23	198	7.9	385	4	US-09-833-381-1646	Sequence 1646, Ap
24	178.5	7.1	293	4	US-09-313-294A-7134	Sequence 7134, Ap
25	140.5	5.6	263	4	US-09-313-294A-686	Sequence 686, App
26	136	5.4	232	4	US-09-833-381-745	Sequence 745, App
27	134.5	5.4	421	4	US-09-833-381-1647	Sequence 1647, Ap
28	123.5	4.9	1077	4	US-09-543-681A-1814	Sequence 1814, Ap
29	123	4.9	4488	4	US-09-543-681A-3962	Sequence 3962, Ap
30	122	4.9	1664976	4	US-08-916-421B-1	Sequence 1, Appli
31	105.5	4.2	3756	4	US-09-813-742A-2	Sequence 2, Appli
32	105.5	4.2	3760	4	US-08-261-663A-3	Sequence 3, Appli
33	105.5	4.2	3760	4	US-09-357-206A-2	Sequence 2, Appli
34	105.5	4.2	3760	5	PCT-US95-07754A-3	Sequence 3, Appli
35	104	4.1	16592	4	US-08-956-171E-53	Sequence 53, Appl
36	103.5	4.1	9510	4	US-09-453-702B-256	Sequence 256, App
37	103	4.1	400	4	US-09-956-171E-4047	Sequence 4047, Ap
38	102.5	4.1	2412	4	US-08-158-232-9	Sequence 9, Appli
39	102.5	4.1	2412	4	US-08-304-626-9	Sequence 9, Appli
40	102.5	4.1	2412	4	US-08-316-301A-11	Sequence 11, Appl
41	102.5	4.1	2412	4	US-08-611-928-9	Sequence 9, Appli
42	102.5	4.1	2412	3	US-09-173-891-9	Sequence 9, Appli
43	102.5	4.1	2412	3	US-09-076-137-11	Sequence 11, Appl
44	102.5	4.1	2412	4	US-09-738-363-11	Sequence 11, Appl
45	102.5	4.1	2412	5	PCT-US92-03624-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-819-993-1
; Sequence 1, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-09-819-993-1

Alignment Scores:
Pred. No.:
Score: 2511.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4

Length: 2002
Matches: 478
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-622-516-2 (1-478) x US-09-819-993-1 (1-2002)

```
Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 40

Qy   21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGlyLysTyrAsp 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  145 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 204
```

RESULT 2
US-10-193-295-1
; Sequence 1, Application US/10193295
; Patent No. 6620608
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL00119SDIV
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Human
US-10-193-295-1

Alignment Scores:
Pred. No.:
Score: 2.6e-302   Length: 2002
                              2511.00   Matches: 478
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%   Indels: 0
DB: 4   Gaps: 0

US-10-622-516-2 (1-478) x US-10-193-295-1 (1-2002)

RESULT 3

```
Qy  121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal 140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  505 GAAGGAATCGACACAACTAATGCAATGTTATGGAGGCACAGCTGCTGTTCAATGCTGTT 564

Qy  141 AsnTrpIleGluSerSerSerTrpAspGlyLeuArgGlyThrHisMetGlnHisAlaTyr 160
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  565 AACTGGATTGAGTCCAGCTCTTGGGATGGGCTTCGTGGGACACATATGCAACATGCCTAT 624

Qy  161 AspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyIleLysLeuSerIle 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  625 GATTTTTACAAGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAAAACTCTCCATTA 684

Qy  181 GlnCysTyrLeuSerAlaLeuLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  685 CAGTGCTACCTCAGTGCATTAGACCGCTGCTACTCTGTCTACTGCAAAAAGATCCATGCC 744

Qy  201 GlnTrpGlnLysGluGlyAspAlaPheThrLeuAsnAspSerPheGlyPheMetIle 220
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  745 CAGTGGCAGAAAGAGGGAGATGCATTTACCCTGAATGATAGTTTTGGGTTCATGATC 804

Qy  221 PheHisSerProTyrCysLeuValGlnLeuLysSerLeuAlaArgMetLeuLeuAsnAsp 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  805 TTTCACTCACCATATTGTAAACTGGTCAGAAAATCTTAGCTCGGATGTTGCTCAATGAC 864

Qy  241 PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyTyrLeuGluAlaPheGly 260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  865 TTCCTTAATGACCAGAATAGAGATAAAAATAGTATTTATAGTGGCTATCTGGAAGCCTTTGG 924

Qy  261 AspValLysLeuGluAspThrThrTyrPheAspArgGlyAspValGlyLysSerAlaPheMetLysAla 280
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  925 GATGTTAAATTAGAAGACACCACTTACTTTGATAGAGGTGGAGAAGGCATTTATGAAGGCT 984

Qy  281 SerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerLeuSerAsnGlnAsnGly 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  985 AGCTCTGAACTCTTCAGTCAGAAAACAAAGGCATCTTTACTTGTATCAAATCAAAATGGA 1044

Qy  301 AsnMetTyrThrSerSerValTyrGlySerLeuValAlaSerValLeuAlaGlnTyrSerPro 320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1045 AATATGTACACATCTTCAGTATATGGTTCCCTTGTAGCATCTGTTCTAGCACAGTACTCACCT 1104

Qy  321 GlnGlnLeuAlaGlyLysLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThr 340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1105 CAGCAATTAGCAGGAAAGAAATGGGAGTGTTTTCTTAGGTTCTGGGTTTGGCGCCACT 1164

Qy  341 LeuTyrSerLeuLysThrGlnGlnProGlySerAlaLeuLeuAspLysLysSerIleThr 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1165 CTGTACTCTCTTAAAGTCACAACAGATGCTACACCGGGTTCTGCTCTTTGAATAAAATAACA 1224

Qy  361 AlaSerLeuCysLysAspLeuLysSerArgGlyThrGlyValLysValAlaProAspAsnVal 380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1225 GCAAGTTTATGTAGATCTTAAATCAAGGCTTGATCTAAGAACTGGTGTGCCACCAGATGTC 1284

Qy  381 PheAlaGluAsnMetLysGlnAspThrHisHisLeuValAsnTyrIleProGln 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1285 TTCGCTGAAAACATGAAGCTCAGAGAGACACCACCATCATTGGTCAACTATATTCCCCAG 1344

Qy  401 GlySerIleAspSerLeuPheGluGlyTyrTrpTyrLeuValArgValAspGluLysHis 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1345 GGTTCAATAGATTCAACTCTTTGAAGGAACGGTGGTACTTAGGTAGGTAGATGAAAAGCAC 1404

Qy  421 ArgGlnThrTyrAlaAlaArgProThrProAsnAspAspThrLeuAspArgGluGlyValGly 440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1405 AGAGAACTTACGGCTCGGGGTCCCACTCCAAATGATGACACTTTGGATAGAGAGTAGGA 1464

Qy  441 LeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysSerValProArg 460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1465 CTTGTGCATTCAAACATAGCAACTGAGCATATTCCAAGCCCTGCCAAGAAAAGTACCAAGA 1524

Qy  461 LeuProAlaThrAlaLeuGluProGluLeuAlaValIleSerArgAsnGlyGluHis 478
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1525 CTCCCTGCCACAGCAGCGCAGAAACCTGAAGCAGCTCTCATTAGTAGTAATGGGGAAACAT 1578
```

US-10-622-516-2 (1-478) x US-08-305-505-1 (1-1824)

```
Alignment Scores:
Pred. No.:
Score:                8.76e-250      Length:        1824
                      2089.00        Matches:       399
Percent Similarity:   83.33%         Conservative:  36
Best Local Similarity: 76.44%        Mismatches:    43
Query Match:          83.19%         Indels:        44
                      1              Gaps:          3
```

```
Qy  1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleLeuVal 20
      |||::||||||::||||||:||||::
Db  1 ATGGCTGGGTCTCTTCCAGTGAACACTGAATCCTGGCCAAAGATGTGGGATTGTT 60

Qy 21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlyIleAlaGluLeuGluLysTyrArgAsp 40
      |||||||||||||||||||||||||||||||||||||||||||||||||
Db 62 GCACTGGAAAATCTATTTCCCTCCCAGTATGTCGACACCAGACTGAGCTGGAAGAGTATGAC 121

Qy 41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr 60
      ||||||||||||||||||||||||||||||||||||||||||||||:
Db 122 GGTGTGGATGCAGGCAAATACACCATTGGGTTAGGCCAGGCAAAAGATGGGCTTCGTGCT 181

Qy 61 AspArgGlyAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
      ||||||||||||||||||:::
Db 182 GACCGAGGAGATATCAATTCCCTCTGTTTGACTGTCGTTCAGAAGCTTATGGAGAGGAAC 241

Qy 81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
      :::||||||||||||||::
Db 242 AGCCTTTCCTATGACTGCAATGGGAGACTGGAAGTTGGAACGGAGACAATAATTGATAAA 301
```

RESULT 4
US-09-819-993-3

; Sequence 3, Application US/09819993
; Patent No. 6436692
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001195
; CURRENT APPLICATION NUMBER: US/09/819,993
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 28001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(28001)
; OTHER INFORMATION: n = A,T,C or G
US-09-819-993-3

Alignment Scores:
Pred. No.:
Score: 5.76e-92
Pred. No.:   839.50   Length: 28001
Percent Similarity: 19.23%   Matches: 302
Percent Identity: 19.16%   Conservative: 1
Best Local Similarity: 33.43%   Mismatches: 3
Query Match: 4   Indels: 1271
DB: Gaps: 6

US-10-622-516-2 (1-478) x US-09-819-993-3 (1-28001)

```
Qy    204 ------------------------------------------------------------ 204
Db  19024 CTGAGGTGTATCTCTGCAAAAATATTTAGGTCGGTTTACCCCTTGTAAGAAAATCAAAGT 19083

Qy    204 ------------------------------------------------------------ 204
Db  19084 GGAGAAAAGAAGGTAAGGTAAGTTGAATTTTGTTCATCTTTTGAGAGAGGTATTTAACAAGGTT 19143

Qy    204 ------------------------------------------------------------ 204
Db  19144 TTGGACTACAGCTGTGATTCAGGGAAAGCTAATGAAAATGAATTACTAAAGTGATCTTAC 19203

Qy    204 ------------------------------------------------------------ 204
Db  19204 CCCAAAAATAATCTTTTTTGCACTTGACCTGTGAATTTGTATTTGTTTTTTTACTGTTATC 19263

Qy    204 ------------------------------------------------------------ 204
Db  19264 ATTAATCTGGAAATTTGTTGAGGCACTGAAAGGACAGTATTTGAGTTAAATGCTATCATAA 19323

Qy    204 ------------------------------------------------------------ 204
Db  19324 CACATTATTACATAAAGTATACTTTTTCTGTAGTCCAACTTTGTTTTTAGAGGTTATGA 19383

Qy    204 ------------------------------------------------------------ 204
Db  19384 GAAGGGGGTTAAAAATCATTATTCAAATGACAAAATATCAGTGAATTTAGTCGCTCTGGATAAG 19443

Qy    204 ------------------------------------------------------------ 204
Db  19444 AAGCAATTCTTGCAGTATATATTAAACAGAATAGTGGTTTTCTAACTTTTTTATTAGGACC 19503

Qy    204 ------------------------------------------------------------ 204
Db  19504 ACAGTAAGAAGTACATGTTACATTGTATGTGTATGCCAGAACTGAAACAAAAAATGTCATGA 19563

Qy    204 ------------------------------------------------------------ 204
Db  19564 CATTACTTACCCTTGCTCGCAAGTTATTCAGTTGCTATTTTTCTACTGCCATTTGTTTTT 19623

Qy    204 ------------------------------------------------------------ 204
Db  19624 TAAAATACTCTTTTATTTAAAAAAATACTAATCCTGACCCACTAAAATTGATTAATGTAAC 19683

Qy    204 ------------------------------------------------------------ 204
Db  19684 CTGCTAAGTGTGTATGAATCTTAAATTTGAAAATTAGTGACATAGTACACATATTGTTTCATC 19743

Qy    204 ------------------------------------------------------------ 204
Db  19744 TTTGAGTGTCTTTTTAAATGTATACTTTTAAGGTATAGAGAGGTTTCATTATACAGTGTAT 19803

Qy    204 ------------------------------------------------------------ 204
Db  19804 TTGTGGTTGCTGTTAAACATATACAAAATATCCTAGCTTATTCTAAAGTCAAACTTTAA 19863

Qy    205 ----------------------------------------------------GluGlyAsnAs 208
Db  19864 AATTTCATGGCTTATATGAATTTCATAGTTTCCTTGGACTTCTCTTTCAGAGGGAAATGA 19923

Qy    208 plysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysLe 228
Db  19924 TAAAGATTTTACCTTGAATGATTTTGGCTTCATGATCTTCACTCCATATTGTAAACT 19983

Qy    228 uValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAs 248
Db  19984 GGTTCAGAAATCCTCAGTCTCGGATGGTTGCTGAATGACTTCCCTTAATGACCAGAATAGAGA 20043

Qy    248 plysAsnSerIleTyrSerGlyLeuGluAlaPheGly 260
Db  20044 TAAAAATAGTATCTATAAGTGGGCCTGGAAGCCCTTTGG-GTAAGAGGAGCTATTATGAGTTT 20102
```

```
Qy    260 ------------------------------------------------------------ 260
Db  20103 TTTCCTTCTATATTAGAGGCATTTTTAATATCTGTTAAGCTGTTTATTTGTACAGAACCTGAG 20162

Qy    260 ------------------------------------------------------------ 260
Db  20163 AAATTGAGAGTCAGAAGAATCTTAGAAGTCATCCAGTCTTAATCTGTGTGTCTCAGTCAGT 20222

Qy    260 ------------------------------------------------------------ 260
Db  20223 GAAGAAATCTAAGTCCAGAGAGGTGGTAGTTAACATGCACAAAATTCTTTAGACATTTCTAT 20282

Qy    260 ------------------------------------------------------------ 260
Db  20283 TCAGAATTTTCTGATTTATTTCTTTCAGCTCCATTCATGTTGTCACGATAAAGTAACTGCA 20342

Qy    260 ------------------------------------------------------------ 260
Db  20343 CAAGGGCTATATTCACTACAGCAGCCTCTTAACTCCTCTCTCAGCACCCCTGCC 20402

Qy    260 ------------------------------------------------------------ 260
Db  20403 CCCATGCCCTTTCCATCCTGCACACTGCCACAGCTAAAGTCAGCTTTGTTGTACTCCACCT 20462

Qy    260 ------------------------------------------------------------ 260
Db  20463 GTCTTTTTCTCACTTAGGCTCCCTAGCAATGCTATGGTGTTCAACTCGGTTCGTTTCTC 20522

Qy    260 ------------------------------------------------------------ 260
Db  20523 CCTGTGTCTCTTGTGTGTCCTTTCTCTATCTGATAAAATTATATACTTGACTTTTAAAACTT 20582

Qy    260 ------------------------------------------------------------ 260
Db  20583 GGCTCCTGTAATACCATGACTTTCTAACTAAAATAAACATTATTATGGACTTGAAATAGT 20642

Qy    260 ------------------------------------------------------------ 260
Db  20643 ATTCTATTCAGTTGATGAATATTCAGTTGAATTCTATTCATTGAAGCCAATATA 20702

Qy    260 ------------------------------------------------------------ 260
Db  20703 AGTGAAATAAATAAAATATAAAAGCTACACAGTGCTCTCTTTAAACCTATTCAAAATCAGGCTTA 20762

Qy    260 ------------------------------------------------------------ 260
Db  20763 ACTTGATTATGAAAACTTTGGAGAAAAAGAACCATATATATACTGTTATGATTTCTA 20822

Qy    260 ------------------------------------------------------------ 260
Db  20823 TAGCAATTAGATTGCTGCTACTTGGCTTTTAATAAATGAGAAAAACAAATTATATACACTTA 20882

Qy    260 ------------------------------------------------------------ 260
Db  20883 AAGATTTGAATCCTRAATTAGGGCCTGCTGTTGTGTAATAAAAACATAGGCTTTAAAACAC 20942

Qy    261 AspValLysLeuGluAspThrTyrPheAspAr 271
Db  20943 TGTAAAACTGTAAAATAAATAAATCTTTCAGGGAATGTTAAATTAGAAGACACCTTGATAG 21002

Qy    271 gAspValGlyLysAlaPheMetLysAlaSerSerGluLeuPheSerGluLeuLysThrLysAl 291
Db  21003 AGATGTGGAGAAAGGCATTATGAAGGCTAGCTCTGAACTCTTCAGTGAGAAAACAAAGGC 21062

Qy    291 aSerIleLeuValSerAsnGlyAsnMetThrThrSerSerValTyrGlySerIle 311
Db  21063 ATCTTTACTTGTATCAAAATCAAAATGGGAAATATGTAACACATCTTCAGTATATGGTTCCCT 21122

Qy    311 uAlaSerValLeuAlaGln- 317
Db  21123 TGCATCTGTTCTAGCACAGTAAGTATAAATTCACCTACTACTTAACTCCCTTATTTGG 21182

Qy    317 ------------------------------------------------------------ 317
```

Db 21183 GAGAGTGTTAGATTTCTTAGAGACCAAAATCTAGTGTCTTAGTGTTGGTGTTGGTAGATCACAGGAA 21242
Qy   317 -- 317

Db 21243 AATTTTATCTTGAGGCTCTCTAATCTGCTTATTGTGTCCATTGACTTGAAAAGATGTTATGGGTT 21302
Qy   317 -- 317

Db 21303 GAGGCTACAGTTCTTCCCAGAAGTGTTTGTTAATTTCATACTGGCTTTCCTGGCTTCTGTT 21362
Qy   317 -- 317

Db 21363 TTCATGGTGTTTTTAATTCTTGACCTACAGTTGAAACCATAAAATACCTGGTTGATGAAGTAA 21422
Qy   317 -- 317

Db 21423 CTTGTTTTTTGTTGGCATGACTTTCACAAGCTCTGTTCATTCCCCACACAAGATGAAAACTCACAT 21482
Qy   317 -- 317

Db 21483 GCTGCAAATATAAAACTAAGTTATATTCCCTACTGCAATAATTAACACTTTGAGTTAGATC 21542
Qy   317 -- 317

Db 21543 CTTAAAAACTTTAAGTTAGTAGATTCTACTTTTACTTATAGCCTAAAATTTTTATTGCTACTTTT 21602
Qy   317 -- 317

Db 21603 ATAGCTTCCCACACGCCTGTAGCTTTGGATCAGTTAAACTTCTGAACTATTGTTTACACCCT 21662
Qy   318 ------TyrSerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySe 335

Db 21663 ACATAGGTACTCACCTCAGCAATTAGCAGGGAAGGAAGTGGTTGTTTTCTTATGGTTC 21722
Qy   335 rGlyLeuValAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrPro------ 352
      |||||||||||||||||||||||||||
Db 21723 TGGTTTGGCTGCCCACTCTGTACTCTCTTAAAGTCACACAAGATGCTACACCGGGTAAGTG 21782

Db 21783 CTGAATCTTTCAACAAGAATGTATTGAGAACTGAGTCCAGGCACAGTGGCTCACACCCGT 21842
Qy   352 -- 352

Db 21843 AATCCCAGCAGTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCGAGACC 21902
Qy   352 -- 352

Db 21903 AGTCTGGCTAACATGGGCTGAAACCCCATCTCTACTAAAAAATACAAAAAATTAGCCAGGTGA 21962
Qy   352 -- 352

Db 21963 GGTGGTGTCCATGCCTGTAGTCCTAGCTACTTGGGAGGCTGAAGTGAGGAGAATCACTTGAAT 22022
Qy   352 -- 352

Db 22023 CCAGGAGGAGGAGGTTGTGGTGAGCCAAGATCACACCACTGTGCTCCAGCCTGGGTGACA 22082
Qy   352 -- 352

Db 22083 GAGCGAGACTCTGTCAAAAAAAAAAAAAATGTATTGAGAACTACTTCTGGGGAGTT 22142
Qy   352 -- 352

Db 22143 GATTTAGCAGTGCTTCTCAAGTGAGCACCTGAATCTGTCCCACAGATCATTACNAATATATTTT 22202
Qy   352 -- 352

Db 22203 AGTCTTCATTACTTCTTTCAGTAGGTTTTTACTCTCTGCCCTAAAAAATCTATCCAAAAAA 22262
Qy   352 -- 352

Db 22263 AAAAAAAAAATTCTACCTTATCTCTGGATAAAGGGATTAGGACTAAGTATTGTCTTAAATTTTTATAGG 22322
Qy   352 -- 352

Db 22323 CTTTATGGTGTCTTGGCTATATTTAAGGTCACTTTTGTGCTTTCCCTGAGCAGGAAAGAGCAA 22382
Qy   352 -- 352

Db 22383 AAATGTAGGAGGATAAAACTGATGAAAAACTTGACATTACTTTTTAAAATTATACCATGGGGCCA 22442
Qy   352 -- 352

Db 22443 GCTGCAAATGGCTCACACCTATAAATCCCAAACACTTCAGGAGGCCGAGGTGGGAGGATTGCT 22502
Qy   352 -- 352

Db 22503 TGAGGCCAGAGTGTTCAAGGCCAACCTGAGCAACATAGTGAGACCCCATCTCTAAAAAA 22562
Qy   352 -- 352

Db 22563 TAATAAAAATAAAAATAATTATACCATGGATGAATTGCTGAGAAGAGTTATTATGTTTCAA 22622
Qy   352 -- 352

Db 22623 ATTATGCCTGTTTCCTAACTTGTCTAGTGGCAGATACTCAATAAATAGATTTCTAGTCTGA 22682
Qy   352 -- 352

Db 22683 CATCATAGGAGAGATTTGTCAAAATAGGTATCACCTTACTCTTTAACTACAATCAGTAGCCAGTA 22742
Qy   352 -- 352

Db 22743 GTTTTAATGAAAATGAAAAGTTGTTTTGCCCTCATTTGGCAACATTTTTACTTAGGCTTCTT 22802
Qy   353 --GlySerA 355

Db 22803 TTGGACATGATTTTTCAAAAAAATCTTTTAAAGTTGAAATTGAAATGGGGTCTG 22862
Qy   355 laLeuAspLysIleThrAlaSerleuCysAspLeuLysSerArgLeuLeuAspSerArgThrG 375
      ||||||
Db 22863 CTCTTGATAAAAATAACACGAAGTTATGTGATCTTAAATCAAGGCTTGATCAAGAACTG 22922

Db 22923 GTGTGGCCACCAGAGTGTCTTCGCTGAAAAACATCAGAGAGCACACCCATCATTTGG 22982
Qy   375 lyValAlaProAspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHis---- 393
      |||||||||||||||
Db 22983 GTAAAAATAATTATTAAAGTTCTTTAAGTTAACCCATTTGGAGGGCTGATATCATTAAGGATG 23042

Db 23043 CTACATATACGGATAAGGATATCAAAGACTTTACTCACTACTAAATCTTGATCTGAAAAT 23102
Qy   393 -- 393

Db 23103 TATTTGGGATATATGAAAACTTATCTTTAGCTTATTTACCAGATGAATTGTATATCATAACT 23162
Qy   394 ---------------------------LeuValAsnTyrIleProGlnGlySerIleAspS 405

Db 23163 AATTGTAGATATTCTCTCCCTTTCCTTTAGTCAACTATATATTTCAACATAAGAGATT 23222
Qy   405 erLeuPheGluGlyThrTrpTyrLeuValArgValAlaAspGluLysHisArgArgThrTyrA 425
      |||||||||||||||
Db 23223 CACTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGATGAAAAACAGCAGAAGAACTTACG 23282

Db 23283 CTCGGGGTTCCCACTCCAAATGATGACACTTGGATGAAAGAGTAGGACTTGTGCCATTCAA 23342
Qy   425 laArgArgProThrProAsnAspAspThrLeuAspGluGlyValGlyLeuValHisSerA 445
      ||||||||||||||||||
Db 23343 ACATAGCAACTGAG----------------------GTAAATAAAAGAGTTCCC 23374
Qy   445 snIleAlaThrGluHisIleProSerProAlaLysLysValPro 459
      ||||:::||||||

RESULT 5
US-10-193-295-3
; Sequence 3, Application US/10193295
; Patent No. 662060B
; GENERAL INFORMATION:
;   APPLICANT: GONG, Fangcheng et al.
;   TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL00119SDIV
;   CURRENT APPLICATION NUMBER: US/10/193,295
;   CURRENT FILING DATE: 2002-07-12
;   PRIOR APPLICATION NUMBER: 08/819,993
;   PRIOR FILING DATE: 2001-03-29
;   NUMBER OF SEQ ID NOS: 5
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 3
;   LENGTH: 28001
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(28001)
;   OTHER INFORMATION: n = A,T,C or G
US-10-193-295-3

Alignment Scores:
Pred. No.:
Score: 5.76e-92
Score: 839.50   Length: 28001
Percent Similarity: 19.23%   Matches: 302
Best Local Similarity: 19.16%   Conservative: 1
Query Match: 33.43%   Mismatches: 3
                                              Indels: 1271
                                              Gaps: 6

US-10-622-516-2 (1-478) x US-10-193-295-3 (1-28001)

```
Qy   150 GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPheTyrLysProAspMetLeuSer 169
Db 18664 GGGCTTCGTGGGGACACAATGCAACATGCCTATGATTTTTACAAGCCTGATATGCTATCT 18723

Qy   170 GluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrIleuSerAlaLeuAspArg 189
Db 18724 GAATATCCTATAGTAGATGGAAAACTCTCCATACAGTGCTACCTCAGTGCATTAGACCGC 18783

Qy   190 CysTyrSerValTyrCysGlyLysIleHisAlaGlnLnTrpGln-Lys-  204
Db 18784 TGCTATTCTGTCCTACTGCAAAAAGATCCATGCCCAGTGGCAGAAGAAGGTAAGTTTTACCCA 18843

Qy   204 ------------------------------------------------------------ 204
Db 18844 TTTTCCTTGGTTTTGGTATGAGTTGAGAGCCAGTCTAAGTACTAGGTATCTTTGGTAGGC 18903

Qy   204 ------------------------------------------------------------ 204
Db 18904 AACTACTTTGTGGGGCATTCTTCATTAATATCCTTTACCATTAATTCCTCACCACCA 18963

Qy   204 ------------------------------------------------------------ 204
Db 18964 ACAACATTTTCCCATAGTTTCTGGGGAAAGTGTAATTACTAGAAGAGGTAAACTTTGGAA 19023

Qy   204 ------------------------------------------------------------ 204
Db 19024 CTGAGGTGTATCTCTGCAAAAAATATTTAGGTCGGTTTACCCCTTGTTAAGAAAATCAAAGT 19083

Qy   204 ------------------------------------------------------------ 204
Db 19084 GGAGAAAAGAAGAAGGTAAGTTGAATTTTGTTCATCTCTTTTGAGAGAGAGGTATTTTAACAAGGTT 19143

Qy   204 ------------------------------------------------------------ 204
Db 19144 TTGGACTACAGCTGTGATTCAGGGGAAAGCTTAATGAAAATGAATTACTAAAGTGATTCTTTAC 19203
```

```
Qy   204 ------------------------------------------------------------ 204
Db 19204 CCCAAAAAATAATCTTTTTGCACTTGACCTGTGAATTGTGTTTTTTTTTTTACTTGTATC 19263

Qy   204 ------------------------------------------------------------ 204
Db 19264 ATTAATCTGGAAATTTGTTGAGGCACTGAAAGGACAGTATTTGAGTTAATGCTATCATAA 19323

Qy   204 ------------------------------------------------------------ 204
Db 19324 CACATTATTACATAAAGTATACTTTTTCTGTAGTCCAACTTGCTTTTTAGAGGTTATGA 19383

Qy   204 ------------------------------------------------------------ 204
Db 19384 GAAGGGGTTAAAAATCATTATTTGGAGGCACTGAAAGGACAGTATTTGAGTTCGCTGGATAAG 19443

Qy   204 ------------------------------------------------------------ 204
Db 19444 AAGCATTCTTGCAGTATATATTAACAGATAGTGGTTTTCTTAACTTTTTTTTTATTAGGA 19503

Qy   204 ------------------------------------------------------------ 204
Db 19504 ACAGTAAGAAGTACATGTTACATTGTGTATGCCAGACTGAAACAAAAAATGTCATGA 19563

Qy   204 ------------------------------------------------------------ 204
Db 19564 CATTACTTACCCTTGCTGCAAGTATTATTCAGTTGCTATTTTTCTACTGCATTTGTTTTT 19623

Qy   204 ------------------------------------------------------------ 204
Db 19624 TAAAATACTCTTTTTATTTAAAAAAATACTAATCCTGACCCACTAAAATTGATTATGTAAC 19683

Qy   204 ------------------------------------------------------------ 204
Db 19684 CTGCTAAGTGTATGAAATCTTAAATTTGAAAAAATTAGTGACATAGTACATATTTGTTTCATC 19743

Qy   204 ------------------------------------------------------------ 204
Db 19744 TTTGAGTGTCTTTTTAAATGTATACTTAAGGTATAGAGAGGTTTCATTATTATACAGTGTAT 19803

Qy   204 ------------------------------------------------------------ 204
Db 19804 TTGTGGTTGCTGTTTAAACATATACAAAATATCCTAGCTTTTGAGTTCAAAACTTTAA 19863

Qy   205 ----------------------------------------------GluGlyAsnAsn 208
Db 19864 AATTTCATGGCCTATATAATGATTTCATAGTTTCCTTGGACTTCTCTTTTCAGGGGAAATGA 19923

Qy   208 pLysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysLysLe 228
Db 19924 TAAAGATTTTTACCCTTGAAATGATTTTGGCCTTCATGATTCTTCACTCACCATAATGTAAACT 19983

Qy   228 uValGlnLysSerLeuAlaArgMetLeuLeuAsnAsnAspPheLeuAsnAsnAspArgAs 248
Db 19984 GGTTCAGAAAATCCTAGCCTCGGATGTTGCTGAAGTGACTTCCTTAATGACCAGAATAGAGA 20043

Qy   248 pLysAsnSerIleTyrSerGlyLeuGluAlaPheGly- 260
Db 20044 TAAAAAATAGTAATCTTATAGTGGCCCTGGAAGCCTTCG-GTAAGAGGAGCTATTATGAGTTT 20102

Qy   260 ------------------------------------------------------------ 260
Db 20103 TTTCCTTCTATATATTAGAGACATTTTAATATCTGTTAAGCTGTTATTTGTACAGACCTGAG 20162

Qy   260 ------------------------------------------------------------ 260
Db 20163 AAATTGAGAGTCAGAAGAAATCTTAGAAGTCATCCAGTCTAAATCTGTGTCTCAGTCAGT 20222

Qy   260 ------------------------------------------------------------ 260
Db 20223 GAAGAATCTAAGTCCAGAGAGAGGTAGTTAACATGCACAAAATTCTTTAGACATTTCTAT 20282
```

```
Db  20283 TCAGATTTTCTGATTTATTTCTTTCAGCTCCATTCATGTTGTCACGATAAAGTAACTGCA 20342
Qy    260 ------------------------------------------------------------   260

Db  20343 CAAGGGCCTATATTCACTACAGCAGGCCTTCTTAACTCCTTCTCAGCACCCCTGCC      20402
Qy    260 ------------------------------------------------------------   260

Db  20403 CCCATGCCCTTTTCCATCCTGCACACTGCCACAGCTAAAGTCAGCTTTTGTACTCCACCT  20462
Qy    260 ------------------------------------------------------------   260

Db  20463 GTCTTTTTCTCACTTTAGGCTCCCCTAGCATGCTAGTGTGTTCAACTCGTTCTTGTTTCTC 20522
Qy    260 ------------------------------------------------------------   260

Db  20523 CCTGTGTCTCTGTGTGTGCCTTTCTCTATCTCGATAAAATATTATACTTTTAAAACTT    20582
Qy    260 ------------------------------------------------------------   260

Db  20583 GGCTCCTGTAATACCATGACTTTTCTAACTAAATAAACATTATTATTGGACTTGAAAATAGT 20642
Qy    260 ------------------------------------------------------------   260

Db  20643 ATTCTATTCAGTTGATGAATATTCAGTTGATTGAATATTCTATTCATTGAAGGCCAATATA 20702
Qy    260 ------------------------------------------------------------   260

Db  20703 AGTGAATAAAATATAAAAGCTACAGTGCGTCGTCTTTTAACCTTATTCAAATCAAGGAGGCTTA 20762
Qy    260 ------------------------------------------------------------   260

Db  20763 ACTTGATTATGAAAACTTTTGAGAAAAAGAACCATATATACAACTGTTATGATTTTCTA   20822
Qy    260 ------------------------------------------------------------   260

Db  20823 TAGCAATTAGATTGCTGCTACTTGGCTTTTAATAAATGAGAAAAACAATTATATACACTTA 20882
Qy    260 ------------------------------------------------------------   260

Db  20883 AAGATTTGAATCCTAATTAGGCCCTGCGTGTTTTAGTGTGTAATAAACATAGGCTTTAAACAC 20942
                                                   AspValLysLeuGluAspThrTyrPheAspAr
Qy    261 ------------------------------------------------------------   271

Db  20943 TGTAAAACTGTAAAATAAATCTTTCAGGGATTGTAAATTAGAAGAACACCTTACTTTGATAG 21002
Qy    271 AGATGTGGAGAAGGCCATTTTATGAAGGGCTAGCTCCTGAACTCTTCTGAACAACAAAGGC 21062
              gAspValGlyLysAlaPheMetLysAlaSerSerGlyLeuPheSerGlyLeuPheSerGlyAl

Db  21003 AGATGTGGAGAAGGCCATTTTATGAAGGGCTAGCTCCTGAACTCTTCTGAACAACAAAGGC 21062
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    291 aSerLeuLeuValSerGlnAsnGlyAsnMetTyrThrSerSerValTyrGlySerIe      311
          ATCTTTACTGTATCAAAATCAAAAATAGGAAAAATGTACACATCTTCAGTATATGGTTCCCT

Db  21063 ATCTTTACTGTATCAAAATCAAAAATAGGAAAAATGTACACATCTTCAGTATATGGTTCCCT 21122
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    311 uAlaSerValLeuAlaGln-                                            317
          TGCACTCTTGTTCTAGCACAGTAAGTATAAATTTCACCTACTTAACTCCCCTTATTGG

Db  21123 TGCACTCTTGTTCTAGCACAGTAAGTATAAATTTCACCTACTTAACTCCCCTTATTGG     21182
Qy    317 ------------------------------------------------------------   317

Db  21183 GAGAGTTAGATTTCTAAGACCAAAATCTAGTGTCAAGCACATGTTGGTGTAGATCACAGAA  21242
Qy    317 ------------------------------------------------------------   317

Db  21243 AATTTTATCTTGAGGCTCTCTAATCTGCTATTGTCCATTGACTTGAAAGATGTAATGGGTT  21302
Qy    317 ------------------------------------------------------------   317

Db  21303 GAGGCTACAGTTCTTCCAGAGAGTATTTGTTAATTCATACTGGCTTTCCTGGCTTCTGTT   21362
Qy    317 ------------------------------------------------------------   317
```

```
Db  21363 TTCATGGTTTTTTTAATTCTTGACCTACAGTTGAACCATAAAATACCTGGTTGATGAAGTAA 21422
Qy    317 ------------------------------------------------------------   317

Db  21423 CTTGTTTTGTGGCATGACTTTCACAAGCTCGTGTTCATTCCCACCAAGATGAAAACTCACAT 21482
Qy    317 ------------------------------------------------------------   317

Db  21483 GCTGCAATATTAAAAACTAAGTTATATATCCCTACTGCAATATTACACTTTGAGTTAGATC  21542
Qy    317 ------------------------------------------------------------   317

Db  21543 CTTAAAACTTTAAGTTAGATTCTACTTTTACTTATAGCCTAAAATTTTATTGCTACTTTT   21602
Qy    317 ------------------------------------------------------------   317

Db  21603 ATAGCTTCCCACACGCCTGTAGCTTTGGATCAGTTAAACTTCTGAACTATTGTTTACACCCT 21662
Qy    317 ------------------------------------------------------------   317

Db  21663 ACATAGGTACTCACCTCAGCAATTAGCAGGGAAGGAAATTGGAGTTGGTTTTCTTATGGTTC 21722
              TyrSerProGlnGlnLeuLeuAlaGlyIleGlyValPheSerTyrGlySe
              |||||||||||||||||||||||||||||||||||||||||
Qy    318                                                                335

Db  21723 TGGTTTGGCTGCCACTCTGTACTCTCTTAAAGTCACACAGATGCTACACCGGGTAAGTG    21782
              rGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrPro-
              ||||||||||||||||||||||||||||||||||||||||||||
Qy    335                                                                352

Db  21783 CTGAATCTTTCAACAAGAATGTATTGAGAACTGAGTCCAGGCACAGTGGCTCACACCCGT   21842
Qy    352 ------------------------------------------------------------   352

Db  21843 AATCCCAGCAGTTTGGGAGGCCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCGAGACC   21902
Qy    352 ------------------------------------------------------------   352

Db  21903 AGTCTGGCTAACATGGCTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGTGA   21962
Qy    352 ------------------------------------------------------------   352

Db  21963 GGTGGGTGCATGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGTAGGAGAATCACTTGAAT  22022
Qy    352 ------------------------------------------------------------   352

Db  22023 CCAGGAGAGGGAGGTTGTGGTGAGCCAAGATCACACACCACTGTGCTCCAGCCTGGGTGACA 22082
Qy    352 ------------------------------------------------------------   352

Db  22083 GAGCGAGACTCTGTCAAAAAAAAAATGTATTGGAAGACTACTCTGGGGAAGTT          22142
Qy    352 ------------------------------------------------------------   352

Db  22143 GATTTAGCAGTCTTCTCAAGTGAGCACCTGAATCGTCCCACAGATCAATTACAATATTTT   22202
Qy    352 ------------------------------------------------------------   352

Db  22203 AGTCTTCATTACTTCTTTTCAGTAGGTTTTTACTCTCTGCCCTAAAAATCTATCCAAAAAA  22262
Qy    352 ------------------------------------------------------------   352

Db  22263 AAAAAAAAAATTCTACCTTATCTGGATAAAAGGATAGGACTAAGTTATCTAATTTTTTATAGG 22322
Qy    352 ------------------------------------------------------------   352

Db  22323 CTTATGGTCTTGGCTATATTTAAGGTCACTTTTGTGCTTTCCCTGAGCAGGAAGAGCAA    22382
Qy    352 ------------------------------------------------------------   352

Db  22383 AAATGTAGAGATAAAACTTGACATTACTTTTTAAAATTATACCATGGGGCA           22442
Qy    352 ------------------------------------------------------------   352

Db  22443 GGTGCAATGGCTCACACCTATAATCCCAACACTTCAGGAGGCTGAGGTGGGAGGATTGCT   22502
```

```
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER APPLICATION FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSEQ for Windows Version 3.0
;
; SEQ ID NO 245
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105)
; OTHER INFORMATION: n=A,T,C or G
US-09-370-838-245

Alignment Scores:
Pred. No.:
Score:                        771.00
Percent Similarity:           99.33%
Best Local Similarity:        99.33%
Query Match:                  30.70%
DB:                           4
```

Alignment Scores:

Score:	2.45e-86
Length:	615
Matches:	148
Conservative:	0
Mismatches:	0
Indels:	1
Gaps:	0

```
US-10-622-516-2 (1-478) x US-09-370-838-245 (1-615)

Qy     3 GlySerLeuProLeuAsnAlaGluAlaAlaCysTrpProLysAspValGlyIleValAlaLeu 22
           | |  | |  | |  | |  | |  | |  | |  | |  | |  | |
Db   108 GGATCACTTCCTTTGAATGCAGAAGCTTGGCTGGCCAAAAGATGTGGGAATTGTTGCCCTT 167

Qy    23 GluIleTyrPheProSerGlnThrTyrValAspGlnHisAlaGluLeuGlyLysTyrArgAspGlyVal 42
Db   168 GAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGGAAAAATAAGATGGTGTA 227

Qy    43 AspAlaGlyLysTyrThrIleGlyTyrLeuGlyGlnAlaLysMetGlyPheCysThrAspArg 62
Db   228 GATGCTGGAAAGTATACCATTGGTTACCTTGGCCAGGCCAAGATGGGCTTCTGCACAGATAGA 287

Qy    63 GluAspIleAsnSerLeuCysMetThrValValGlnIleAsnLeuMetGluArgAsnAsnLeu 82
Db   288 GAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGATCAATTTAATGGAGAGAAATAACCTT 347

Qy    83 SerTyrArgAspCysIleGlyTyrArgLeuGluValGlyThrGluThrIleIleLeuAspLysLys 102
Db   348 TCCTATAGATGCATTGGGTACCGGCTGGAAGTTGGAACAGAGACAATCATCGACAAAAAG 407

Qy   103 SerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGlyGly 122
Db   408 TCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATAGAAGGA 467

Qy   123 IleAspThrThrAsnAlaCysTyrGlyLysTyrThrAlaAlaValPheAsnAla-ValAsnThr 142
Db   468 ATCGACACAACTAATGCATGCTATGGAAGGCAGAGCTGCGTCTTCAATGCTTGTTAAACTG 527

Qy   142 pIleGluSerSerSerTrpAspGly 150
Db   528 GATTGAGTCCAGCTCTTTGGGGATGGA 552
```

```
RESULT 7
US-09-370-838-128
; Sequence 128, Application US/0937083B
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
; CURRENT APPLICATION NUMBER: US/09/370,838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER APPLICATION FILING DATE: 1999-08-09
```

```
Qy   352 -------------------------------------------------------------- 352
Db 22503 TCAGGCCAGATGTTCAAGGCCCAACCTGAGCAACATAGTGAGACCCACTCTATAAAAA 22562

Qy   352 -------------------------------------------------------------- 352
Db 22563 TAATAAAATAAAATAATTATACCATGGATTAATTGTAGACAAGTTATTTATAGTTCAA 22622

Qy   352 -------------------------------------------------------------- 352
Db 22623 ATTATGCCTGTTTCCTAACTTGTCTAGTGGGCAGATACTCCAATAATAGATTTCTGA 22682

Qy   352 -------------------------------------------------------------- 352
Db 22683 CATCATAGGAGATTTGTCAAATAGGTACATCTTTATCTTTTAACTACTAGCCAGTA 22742

Qy   352 -------------------------------------------------------------- 352
Db 22743 GTTTTAATGAAAATGAAAAGTTGTTTGCCTCATTGGCAACATTTTACTTAGGCTTCTT 22802

Qy   353 ------------------------------------------------GlySerA 355
                                                            | | |
Db 22803 TTGGACATGATTTTTCAAAAAAATCTTTTAATGTTGAATTATTCACTATTTTAGGGTCTG 22862

Qy   355 1aLeuAspGlyLysIleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThr 375
           |  |  |  |  |  |  |
Db 22863 CTCTTGATAAAATAAACAGCAAGTTTATGTGATCTTAAATCAAGGCTTGATTCAAGAACTG 22922

Qy   375 lyValAlaProAspValPheAlaLeuGlyAsnMetLysGluLeuArgGluAspThrHisHis 393
Db 22923 GTGTGGCACCAGAGTCTTCGCTGAAAAACATGAAGCTCAGAGAGGACACCCATCATTTGG 22982

Qy   393 -------------------------------------------------------------- 393
Db 22983 GTAAAAATATTAAATGTTCTTTAAGTAACCCATTTGGAGGGGGCTGATATCATTAAGGATG 23042

Qy   393 -------------------------------------------------------------- 393
Db 23043 CTACATATACGATAAGGATATCAAGACACTTACTCAGTACTAATCTGATGGTCACTGAAAT 23102

Qy   393 -------------------------------------------------------------- 393
Db 23103 TATTGGGATATATGAAACTTATCTTAGCTTTATTACCAGATGAATTGTATATCATAACT 23162

Qy   393 -------------------------------------------------------------- 393
Db 23163 AATTGTAGATATTCTCTCCCTTTCCTTTAGTCAACTATAATTCCCCAGGGGTTCAATAGAT 23222

Qy   394 -LeuValAsnTyrIleProGlnGlySerIleAspS 405
Db 23223 CACTCTTCTTTGAAGGAACGTGGTACTTAGTTAGGGTGGGATGAAAAGCACAGAGAACTTACG 23282

Qy   405 erLeuProGluGlyThrTrpTyrLeuValValArgValAspSerGluTyrHisArgArgThrTyrA 425
Db 23283 CTCGGGCGTCCCACTCCAAATGATGACACTTTGGATGAAGGAGTAGGAACTTGTCGCATTCAA 23342

Qy   425 1aArgArgProThrProAsnAspAspThrLeuAspProGlyLeuValGlyLeuValHisSerArA 445
Db 23343 ACATAGCAACTGAG-------------------GTAAATAAAAGAGTTCCC 23374

Qy   445 serIleAlaThrGlyHisIleProSerProAlaLysLysValPro 459
```

```
RESULT 6
US-09-370-838-245
; Sequence 245, Application US/0937083B
; Patent No. 644425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.475C1
```

; Alignment Scores:

US-09-370-838-128
; TYPE: DNA
; LENGTH: 500
; ORGANISM: Homo sapien
;
; SEQ ID NO 128
; SOFTWARE: FastSEQ for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 289
; EARLIER FILING DATE: 1999-04-02

Alignment Scores:

Pred. No.:
Score: 5.09e-78
                            703.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 28.00%
DB: 4

Length: 500
Matches: 133
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-622-516-2 (1-478) x US-09-370-838-128 (1-500)

```
Qy     1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 161

Qy    21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   162 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 221

Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyTyrLeuGlnAlaIleAlaMetGlyPheCysThr 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   222 GGTGTAGATGCTGGAAAGTATACCAATTGGGCTTGGGCCCAGGCCAAGATGGGCTTCTGCACA 281

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   282 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 341

Qy    81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   342 AACCTTTCCTATGATTGCATTGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAA 401

Qy   101 SerLysSerValLysThrAsnLeuMetGlnIleuPheGluGluSerGlyAsnThrAspIle 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   402 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 461

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThr 133
           |||||||||||||||||||||||||||||||||||||
Db   462 GAAGGAAATCGACACAACTAATGCATGCTTATGGAGGCACA 500
```

RESULT 8
; US-09-401-064-187
; Sequence 187, Application US/09401064
; Patent. No. 6623923
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeline Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C2
; CURRENT APPLICATION NUMBER: US/09/401,064
; CURRENT FILING DATE: 1999-09-22
; NUMBER OF SEQ ID NOS: 371
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 187
;
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapien
;
; US-09-401-064-187

Alignment Scores:

Pred. No.:   6.84e-76
Score: 686.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 27.32%
DB: 4

Length: 506
Matches: 130
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-622-516-2 (1-478) x US-09-401-064-187 (1-506)

```
Qy     1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 ATGCCTGGATCACTTCCTTTGAATGCAGAAGCTTGCTGGCCAAAAGATGTGGGAATTGTT 176

Qy    21 AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp 40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 GCCCTTGAGATCTATTTTCCTTCTCAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 236

Qy    41 GlyValAspAlaGlyLysTyrThrIleGlyTyrLeuGlnAlaIleAlaMetGlyPheCysThr 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 GGTGTAGATGCTGGAAAGTATACCAATTGGGCTTGGGCCCAGGCCAAGATGGGCTTCTGCACA 296

Qy    61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn 80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GATAGAGAAGATATTAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 356

Qy    81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys 100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAA 416

Qy   101 SerLysSerValLysThrAsnLeuMetGlnIleuPheGluGluSerGlyAsnThrAspIle 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGAAATACAGATATA 476

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyTyr 130
           |||||||||||||||||||||||||||||||||||||
Db   477 GAAGGAAATCGACACAACTAATGCATGCTTAT 506
```

RESULT 9
; Sequence 1, Application US/09306595C
; Patent No. 6284506
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE: ISOPRENOID PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/306,595C
; CURRENT FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 98108210
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron
; LOCATION: (1523)..(1699)
; NAME/KEY: exon
; LOCATION: (1700)..(1826)
; NAME/KEY: intron
; LOCATION: (1827)..(1920)

```
; NAME/KEY:   exon      LOCATION: (1921)..(2277)
; NAME/KEY:   intron    LOCATION: (2278)..(2351)
; NAME/KEY:   exon      LOCATION: (2352)..(2409)
; NAME/KEY:   intron    LOCATION: (2410)..(2497)
; NAME/KEY:   exon      LOCATION: (2498)..(2504)
; NAME/KEY:   intron    LOCATION: (2505)..(2586)
; NAME/KEY:   exon      LOCATION: (2587)..(2768)
; NAME/KEY:   intron    LOCATION: (2769)..(2851)
; NAME/KEY:   exon      LOCATION: (2852)..(2891)
; NAME/KEY:   intron    LOCATION: (2892)..(2985)
; NAME/KEY:   exon      LOCATION: (2986)..(3240)
; NAME/KEY:   intron    LOCATION: (3241)..(3325)
; NAME/KEY:   exon      LOCATION: (3326)..(3493)
; NAME/KEY:   intron    LOCATION: (3494)..(3601)
; NAME/KEY:   exon      LOCATION: (3602)..(3768)
; NAME/KEY:   polyA_site LOCATION: (4043)..(4044)
; LOCATION:   US-09-306-595C-1
```

```
Alignment Scores:
  Pred. No.:

  Score:                    1.36e-61
  Percent Similarity:       585.00
  Best Local Similarity:    37.15%
  Query Match:              25.56%
  DB:                       25.41%
                            23.30%

                 Length:      4775
                 Matches:      181
                 Conservative:  82
                 Mismatches:   129
                 Indels:       318
                 Gaps:          15
```

US-10-622-516-2 (1-478) x US-09-306-595C-1 (1-4775)

```
Qy    25  TyrPheProSerGln-------------------------TyrValAspGlnAla---------  34
Db  1652  TATCTCCCGACGGCGAAATACACACTGACCGCGGATTTCTTCCGATCAGGCCATGCCTCAC 1711

Qy    35  --GluLeuGlyLysTyrAspGlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGln--  53
          :::|||||||||||||||||||||
Db  1712  AAGGATCTCGAGGCTTTTGATGGGGGTTCCTTCCGAAAGTACACCATCGGTCTCGGCAAC 1771

Qy    54  AlaLysMetGlyPheCysThrAspAlaGluAspIleAsnSerLeuCysMet--ThrVal--  72
          |||||     |||                          :::  ||||||
Db  1772  AACTTCATGGGCTTCACGACGACATCAACTCGTTCGGCCTTGACGGGTCAG          1831

Qy    72  -----------------------------------------------------------   72
Db  1832  TCTCTTCCGTTCAGCAATCGACGAGGAAAAGGCCCAAGCGCATCTCACTGACACCTTTC 1891

Qy    73  -------------------------ValGlnAsnLeuMetGluArgAsnAsnLe-------  82
                                   |::                    ||::
Db  1892  TCCGTTTTGCAAATTCCAATTCATTGATTGTTAGCTGTTTCCGTCTCTCTATCACAAACGT 1951

Qy    82  uSerTyrAspCysIleGlyArgLeuGlyTyrThrGluThrIleIleIleAspLysSerIly 102
          :||||||||||                       |:::|||          |:::
Db  1952  TGATCCCAAGTCAAATCGGTCGAATTGATCGTCGAACTGCAGTTGCTACTTTGGACAGTCCAA 2011

Qy   102  sSerValLysThrAsnLeuMetGlnLeuPheGluGlyLeuSerGlyAsnThrAspIleLuGl 122
          |||||||||                             :::  |||
Db  2012  ATCTGTCAAGACAGCAGTCCTTATGGACTTGTTCGAGTCCCACGGGCAACACAGATTGAGG 2071

Qy   122  yIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTr  142
          :::                                       :::
Db  2072  TATCGGACTCCAAGAATGCCTTGCTACGGTTCTACCCGGTTCAAATGCCGTCAAACTG 2131
```

```
Qy   142  pIleGluSerSerTrpAspGlyLeuArg------------------------------  152
          |||||||||||||||
Db  2132  GATCGGAGTCATCCCTCTTTGGGGACGGGAAG-AAATGCCATTGTCTTCTGCGGAGACAATTGCCA 2190

Qy   152  -----------------------------------------------------------  152
Db  2191  TCTACGGCCGAGGGGTGCTGCCGGACCTGCCCGGAGGGTGCTGGTGGCCATCCTCATCG 2250

Qy   152  -----------------------------------------------------------  152
Db  2251  GACCCGGACGGCTCCCGTCCGTCGTTCGAGCGGTGAGTTCCAAATCGGTTCTTCCACGGCA 2310

Qy   153  ----------------------------------------------GlyThrHisMe  156
                                                        ||
Db  2311  GCGGCTGAAACAAACCCTTATCCGTCATTCTCATCAATCTAGCCGTCCACGAAACTTCAT 2370

Qy   156  tGlnHisAlaTyrAspPheTyrLysProAspMetLeuSer-----------------  169
          ||::||:|||||||                 ||:::::
Db  2371  GACCAACGCTTGGGACTTCTTACACAAGCCTAATCTTCTTCGTATGTTCAAATTTTGAAGTT 2430

Qy   169  -----------------------------------------------------------  169
Db  2431  TGGCGCTTGGGAGAGTCTTACACTAATTCGGGGGTGTCGTATCGTTTGTGC 2490

Qy   169  -----------------------------------------------------------  169
Db  2491  TTTATAGTGAATACGTTCGTTGTGCCGCACCTTCCTATATTTTAGTTTTTGATCAAATATTGTC 2550

Qy   170  ----------------------GluTyrProIleValAspGlyLysLeuSe  179
                                 ||||||||||||||
Db  2551  CATTGAAATTAACCTTGAAACCTTCTCCTCCCAAATAG-CCCATTGTCGATGGACCTCTCTC 2609

Qy   179  rIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysArgIleHi 199
          :|||::||||||      ||::||||||
Db  2610  CGTCACTTCCTTACGTCACACGGCGTCGTTCGAACGGCCTTATGAGAGCTTACCGGAACAAAGTATGC 2669

Qy   199  sAlaGlnTrpGlnInLysGluGlyAsnAspLys---------------------------  209
          :|||||                   ||
Db  2670  CAAGGCGATTT----GGAGGACCCAAGAACTAACGGTGTCACCAACGGACACACGGA 2720

Qy   210  -----------------AspPheThrLeuAsnAspPheGlyPheMetIlePheHis----  222
                           :::      ||:::||
Db  2721  GGTTGCCGGTGTCAGTGCGTCGCGTCGTTCGAATTACCTTTGTTCCACAGGTAAGGCTCATC 2780

Qy   223  ------------------------------------------SerProTyr-----  225
                                                    |||
Db  2781  TTCTGTATTCTCCTTAAATTCAACGGATCAACGGAGTTAATTCGGTTCATCATTATATCT 2840

Qy   226  -----------CysGlyLeuValGlnLysSerLeuAlaArgMet-------------  236
                     :::             |||
Db  2841  TGTTCGGAACAGTCCTTACGGAAAAGCAGGTTGTCAAAAGGCCCACGGCGACTTGTTAAGCAGT 2900

Qy   236  -----------------------------------------------------------  236
Db  2901  CTTTTTTGTAACTCTTAGCTTGCAGATAAAAAACTTTAGGTTTGTGGTACTTCATTATTTAT 2960

Qy   237  ----------------------------------LeuLeuAsnAspGlnAsnArgAs  248
                                            |:::            ||:::
Db  2961  GCATCTCTTGAATCACCTTATCTCAGTTGTAGTACAATGACTTCCGAAAACAAACCCCAAC- 3015

Qy   248  pLysAsnSerIleTyrSerGlyLeuGluAla--PheGlyAspValLysLeuGluAspTh 267
          |||                             |:::
Db  3016  ----GACCCGGTTTTTTGCTAAGGTCCCAGCCGACGTTGCTACTTTGGACATGAAGAAAAG 3071

Qy   267  rTyrPheAspArgAspValGlyLysAlaPheMetLysAlaSerSerGluLeuPheSerGl 287
          ||||                       |||
Db  3072  TCTTTCAGACAAGAAATGCGGAGAAATCTCGATGCTAAGTGTCTTCTTCAACA 3131

Qy   287  nLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThrSerSerVa 307
          :::
Db  3132
```

GENERAL INFORMATION:

; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: OJIMA, Kazuyuki
; APPLICANT: SETOGUCHI, Yutaka
; TITLE OF INVENTION: ISOPRENOID PRODUCTION
; FILE REFERENCE: ISOPRENOID PRODUCTION
; CURRENT APPLICATION NUMBER: US/09/925,388
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/306,595
; PRIOR FILING DATE: 1999-05-06
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1
; LENGTH: 4775
; TYPE: DNA
; ORGANISM: Phaffia rhodozyma
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1239)..(1240)
; OTHER INFORMATION: EXPERIMENTAL
; NAME/KEY: exon
; LOCATION: (1305)..(1361)
; NAME/KEY: intron
; LOCATION: (1362)..(1504)
; NAME/KEY: exon
; LOCATION: (1505)..(1522)
; NAME/KEY: intron

RESULT 10
US-09-388-1
Sequence 1, Application US/09925388
; Patent No. 6586202

US-10-622-516-2 (1-478) x US-09-925-388-1 (1-4775)

Alignment Scores:
Pred. No.:
Score: 1.366-61   Length: 4775
Percent Similarity: 585.00   Matches: 181
Best Local Similarity: 37.15%   Conservative: 82
Query Match: 25.56%   Mismatches: 129
DB: 23.30%   Indels: 318
                                4   Gaps: 15

Qy   25 TyrPheProSerGln--------------------------TyrValAspGlnAla--------- 34
Db   1652 TATCTCCGACGCGAAATACAACACTGACCGGCGGATTTCTCTCGAATCAGGCCATCGGCTCAC 1711

Qy   35 ---GluLeuGlyLysTyrAspGlyValAspAlaGlyValLysTyrThrIleGlyLeuGlyGln 53
Db   1712 AAGGATCTCGAGGGCTGTTTTGATGGGGTTCCTTCCGGAAAAGTACACCATCGGTCTCGGCAAC 1771

Qy   54 AlaLysMetGlyPheCysThrAspArgGluAspIleAsnSerLeuCysMet-ThrVal--   72
Db   1772 AACTTCATGGCCTTCACCGACGACGACTGAGGACATCAACTCGTTCGCCGTTTGAACGGTCAG 1831

Qy   72 --   72
Db   1832 TCTCTCTCCGTTTCAGCAATCGACAGGAAAAAGGCCCAAGCGCCATCTCACTGACACCTTTC 1891

Qy   73 ---------------------------ValGlnAsnLeuMetGluArgAsnAsnIle   82
Db   1892 TCCGTTTTGCAATTCCATTTGATTGTTAGCTGTTCCGGTGTCTTCTATCAAAGTACAACGT 1951

Qy   82 uSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerIly 102
Db   1952 TGATCCCAAGTCAATCGGTCGAATTGATGTCGGAACTAGAGTCCAATCCAATCATTGACAAGTCCAA 2011

Qy   307 lTyrGlySerLeuAlaSerValLeuAlaGlnThrTyrSerProGlnGlnLeu---------- 323
Db   3132 GCAGGTTGAGCCTGGAATGACCACCGTCCGACGCTCGGGAAACTTGTACACCGCCTCTCT 3191

Qy   307 lTyrGlySerLeuAlaSerValLeuAlaGlnThrTyrSerProGlnGlnLeu---------- 323
Db   3192 CTTCGGTGCTCCTCGGCAAGTTTGTTCTCTAATGTTCCTGGTGACGAGCTCGTAAGTCTTGA 3251

Qy   323 -- 323
Db   3252 TCTCTATCCAATCATCTCTTCCTTATCAATTGAACTGAACTCTTTTCTTTAATGCTGGC 3311

Qy   324 ------------AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuuAlaA   339
Db   3312 TTTTCTTGAACAGGTCGGCAAGCGCAATTGCTCTCTCTACGGCCTACGGATCTGGAGCTGCTG 3371

Qy   339 laThrLeuThrTyrSerLeuuLysValThrGlnAspAlaThrProGlySerAlaLeuuAspLysI 359
Db   3372 CTTCTTTTCTATTGCTTCTTAAGGTCAAGAGCTCAAAACCGCT------------------T 3410

Qy   359 leThrAlaSerLeuCysAspLeuLysSerArgLeuuAspSerArgThrGlyValAlaProA 379
Db   3411 TCATCTCTGAGAAGCTTGAATCTCAACAACCGATTGAGCAACATGAAGAATTGTCCCGTGTG 3470

Qy   379 spValPheAlaGluAsnMetLyβ------------------------------------ 386
Db   3471 ATGACTTTGTTCAAAAGCTCTGAAGGTACGTTGGAATAATGACTTTTTTTGTGGACCGTGGTC 3530

Qy   386 -- 386
Db   3531 TTTGTCAACCGCTAACAACCTTCTTGAAATCGGTCTCTCTTTTTGGTTTTGAAATTCGCTCGGCG 3590

Qy   387 --------LeuArgGluAspThrHisHisLeuuValAsnTyrIlePtoGlnGlySerI 403
Db   3591 CTTCGACACAGGTCCGAGAAGAGACTCACAACGCCGGTGTCATTATTCGCCCATCGGGTTCGC 3650

Qy   403 leAspSerLeuuPheGluGlyThrTrpTyrLeuuValArgValAspGluLysHisArgArgT 423
Db   3651 TTGACGGATCTTCGGCCTGGATACGTTACTACTTGGGAGAGATTGACCAGCATGTGGGGGTCGAC 3710

Qy   423 hrTyrAlaArgArgProThr 429
Db   3711 AGTACAAGCAGGTCCCTTCT 3730

```
Qy   102  sSerValLysThrAsnLeuMetGlnLeuPheGluSerGlyAsnThrAspIleGluGl   122
          ATCGTCAAGACAACAGTCCTTATGGAACTTCGAGTCCGGAAATTCTGATTGAGGG
          |||||   |||||||||||||| |||||||||||| ||||||||||||||  ||||
Db   2012 ATCGTCAAGACAACAGTCCTTATGGAACTTCGAGTCCGGAAATTCTGATTGAGGG   2071

Qy   122  yIleAspThrThrAsnAlaCysTyrGlyCysGlyThrAlaAlaValPheAsnAlaValAsnTr   142
          TATCGACTACCAAGAATGCCTGCTACGGTTGTGGTTCTGCCGCGGTGTTCAATGCCGTCAACTG
          |||||| |  ::  |||||||  |:|||||||||  ||| || |||||||||||||||||||
Db   2072 TATCGACTACCAAGAATGCCTGCTACGGTTGTGGTTCTGCCGCGGTGTTCAATGCCGTCAACTG   2131

Qy   142  pIleGluSerSerSerThrAspGlyLeuArg                                 152
          GATCGAGTCATCCTCTTGGGACGGAAG-AAATGCCATTGTGTCTTCGCGGAGACATTGCCA
          ||||                        :::
Db   2132 GATCGAGTCATCCTCTTGGGACGGAAG-AAATGCCATTGTGTCTTCGCGGAGACATTGCCA   2190

Qy   152  -------------------------------------------------------------   152
          GACCCGACGGCTCCCGTGTCGTCTTCGAGTTCCAATCCGGTCATTTTCTTCCACGGCA
Db   2191 GACCCGACGGCTCCCGTGTCGTCTTCGAGTTCCAATCCGGTCATTTTCTTCCACGGCA   2250

Qy   152  -------------------------------------------------------------   152
          TCTTACGCCGAGGGTGCTGCCCGGACCTGCCGGAGGCGGTGCCGGCCGCCATCCTCATCG
Db   2251 TCTTACGCCGAGGGTGCTGCCCGGACCTGCCGGAGGCGGTGCCGGCCGCCATCCTCATCG   2310

Qy   152  -------------------------------------------------------------   152
          GCGGCTGAAACAACCCTTATCCGTCATTCTCATCAATCTAGCCGTCCACGGAAACTTCAT
Db   2311 GCGGCTGAAACAACCCTTATCCGTCATTCTCATCAATCTAGCCGTCCACGGAAACTTCAT   2370

Qy   153  ----------------------------------------------GlyThrHisMe   156
          GACCAACGCCTGGGACTTCTACAAGCCTAATCTTTCTTCGTATGTTCAAATTTTGAAGTT
          |||:::||
Db   2371 GACCAACGCCTGGGACTTCTACAAGCCTAATCTTTCTTCGTATGTTCAAATTTTGAAGTT   2430

Qy   156  tGlnHisAlaLeuTyrArgAspPheTyrLysProAspMetLeuSer                 169
          TTTATAGTGAATACGTTCGTCTCGCGACACCTCCTTATATTAGTTTTTTGATCAAATATTGTC
Db   2431 TTTATAGTGAATACGTTCGTCTCGCGACACCTCCTTATATTAGTTTTTTGATCAAATATTGTC   2490

Qy   169  -----------------------------------------------------------    169
          CATTGAATTAACCTGAAACCTTCTCCTCCCAAATAG-----GluTyrProIleValLeu   179
Db   2491 CATTGAATTAACCTGAAACCTTCTCCTCCCAAATAG-----CCCATTGTCGATGGACCTCTC   2550

Qy   179  rIleGlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHi   199
          CCTCACTTCCTACGTCACAACGCCATTGACAAGGGCTTATGAAGCTTACCGAACAAAGTATGC
Db   2551 CCTCACTTCCTACGTCACAACGCCATTGACAAGGGCTTATGAAGCTTACCGAACAAAGTATGC   2609

Qy   199  sAlaGlnInTrpGlnGlyAsnAspLys                                     209
          GGAGGACCCAAGAGACTAAACGGTGTCACCAACGGACGACACCGA
Db   2610 GGAGGACCCAAGAGACTAAACGGTGTCACCAACGGACGACACCGA   2669

Qy   209  -----------------------------------------------------------    209
          CAAGCGATTT-------hrIleTyrAlaArgArgProThr   429
Db   2670 CAAGCGATTT----------------------------------------   2720

Qy   210  -AspPheThrLeuAsnAspPheGlyPheMetIlePheHis                       222
          GGTTGCCGCGTGTCAGTCGCGTTCGATTACCTTTTGTTCCACAGGTAAGGGTCATC
Db   2721 GGTTGCCGCGTGTCAGTCGCGTTCGATTACCTTTTGTTCCACAGGTAAGGGTCATC   2780

Qy   223  ----------------------------------------SerProTyr----   225
          TTCTGTATTCTCCTTAAATTCAACCGATCAACGGAGTTAATTCGTGTCATCATATTATCT
Db   2781 TTCTGTATTCTCCTTAAATTCAACCGATCAACGGAGTTAATTCGTGTCATCATATTATCT   2840

Qy   226  -----------CysLysValGlnLysSerLeuAlaArgMet-------               236
          TGTTGGAACAGTCCTTACGGAAAGCCACCGGACCGACTTGTAAGCAGT
Db   2841 TGTTGGAACAGTCCTTACGGAAAGCCACCGGACCGACTTGTAAGCAGT   2900

Qy   236  -----------------------------------------------------------    236
          CTTTTTGTTAACTCTTAGCTTGCAGATAAAAACTTTAGTTTCTGGTACTCATTATTTAT
Db   2901 CTTTTTGTTAACTCTTAGCTTGCAGATAAAAACTTTAGTTTCTGGTACTCATTATTTAT   2960

Qy   237  --------------LeuLeuAsnAspPheLeuAsnAspGlnAsnArgAs   248
          GCATCTCTTGAATCACCTTATCTAGTGTTACAATGACTTCCGAAACAACCCCAAC-----
          |||||| ::||| ||:::
Db   2961 GCATCTCTTGAATCACCTTATCTAGTGTTACAATGACTTCCGAAACAACCCCAAC-----   3015

Qy   248  pLysAsnSerIleTyrSerGlyLeuGluAla--                              267
          ------LeuLeuAsnAspPheMetLysAlaSerSerGluLeuPheSerGl
Db   3016 --GACCCGGTTTTGTGAAGGTGCCAGCCGACTTGTACTTTGTGCACATGATGAAGAAAAG   3071

Qy   267  rTyrPheAspArgAspValGlyLysAlaPheMetLysAlaSerSerGluLeuPheSerGl   287
          TCTTTCAGACAAGAATGTCGGAGAAATCTCTGATTGCTGCCTCCAAGTCTTCTTCAACAA
          |||||::|        |||||||::| :::      |||||    |||    ::::
Db   3072 TCTTTCAGACAAGAATGTCGGAGAAATCTCTGATTGCTGCCTCCAAGTCTTCTTCAACAA   3131

Qy   287  nLysThrLysAlaSerLeuLeuValSerAsnGlyAsnGlyAsnMetTyrThrSerSerVa   307
          GCAGGTTGAGCCTGGAATGACCAGACCACGTCCGACACGTCGGAAACTTGTACACCGCCTCT
          ::::           :::                 ::||| :::::   |||:::|||
Db   3132 GCAGGTTGAGCCTGGAATGACCAGACCACGTCCGACACGTCGGAAACTTGTACACCGCCTCT   3191

Qy   307  lTyrGlySerLeuAlaSerValLeuAlaGlnTyrSerProGlnGlnLeu--------     323
          CTTCGGGTGCTCCGCAAGTTTGTTCTCTAATGTTCCCTGGTGACGAGCTCGTAAGTCTTGA
          ::: :::                                         :::|||
Db   3192 CTTCGGGTGCTCCGCAAGTTTGTTCTCTAATGTTCCCTGGTGACGAGCTCGTAAGTCTTGA   3251

Qy   323  -----------------------------------------------------------    323
          TCTCTATCCCAATCATCTCCTCTTATCAATTGAACTGAAGCTCTTTTCTTAAATGCTGGC
Db   3252 TCTCTATCCCAATCATCTCCTCTTATCAATTGAACTGAAGCTCTTTTCTTAAATGCTGGC   3311

Qy   324  -----------AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaA   339
          TTTCTCTTGAACAGGTCGGCGAAGGCCATTGCTCTCAACGCCCATCCCATCCGATTCGGC
Db   3312 TTTCTCTTGAACAGGTCGGCGAAGGCCATTGCTCTCAACGCCCATCCCATCCGATTCGGC   3371

Qy   339  laThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysI   359
          TTTCTTGAACAGGTCGGGTGGTCTGGAAGGCTTGTCTCTTTAAGGTCAAGAGACTCAACGCGT
Db   3372 TTTCTTGAACAGGTCGGGTGGTCTGGAAGGCTTGTCTCTTTAAGGTCAAGAGACTCAACGCGT   3410

Qy   359  leThrAlaSerLeuCysTyrAspLeuSerSerArgLeuAspSerArgThrGlyValAlaProA   379
          ATGACTTTGTCAAAGCTCTGAAGGTACGTTGGATAATGACTTTTTTGTGGGATCCGTGGTC
Db   3411 ATGACTTTGTCAAAGCTCTGAAGGTACGTTGGATAATGACTTTTTTGTGGGATCCGTGGTC   3470

Qy   379  spValPheAlaGluAsnMetLys-------------------------------------    386
          TCATCTCGAACAGGTCGGCGAAGGCTTGATCTCAACAACGCGTCGTCATAATCGCCATCC
Db   3471 TCATCTCGAACAGGTCGGCGAAGGCTTGATCTCAACAACGCGTCGTCATAATCGCCATCC   3530

Qy   386  -----------------------------------------------------------    386
          TTGACGATCTCTGGGCCTGGATCGTACTACTTGGGAGAGGATGACAGCACGTGGGGTCGAC
Db   3531 TTGACGATCTCTGGGCCTGGATCGTACTACTTGGGAGAGGATGACAGCACGTGGGGTCGAC   3590

Qy   386  -----------------------------------------------------------    386
          TTTGTCACCGGCTAACAACCTTCTTGAATCGGTCTCTCTTGTTGAAATTCGGCTCGGGCG
Db   3591 TTTGTCACCGGCTAACAACCTTCTTGAATCGGTCTCTCTTGTTGAAATTCGGCTCGGGCG   3650

Qy   387  ----------LeuArgGluAspThrHisHisLeuValAsnTyrIleProGlnGlySerI   403
          TTGACGAATCTCTGGGCCTGGATCGTACTACTTGGGAGAGGATGACAGCACGTGGGGTCGGAC
Db   3651 TTGACGAATCTCTGGGCCTGGATCGTACTACTTGGGAGAGGATGACAGCACGTGGGGTCGGAC   3710

Qy   403  leAspSerLeuPheGluGlyTyrThrArgTyrLeuValArgValAspGluLysHisArgT   423
          CTTCGGACACAGGTCGGCGAAGGAGAAGACGCGTGTCATAATCGGCCATCGGTTCGC
Db   3711 AGTACAAGACAGGTCCCTTCT   3730

Qy   423  hrTyrAlaArgArgProThr   429
          ATGACTTTGTCAAAGCTCTGAAGGTACGTTGG
```

```
SOFTWARE:  PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
;   APPLICATION NUMBER:  US/08/998,416
;   FILING DATE:  24-DEC-1997
;   CLASSIFICATION:
PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:  435
;   FILING DATE:  31-DEC-1996
ATTORNEY/AGENT INFORMATION:
;   NAME:  Meigs, J. Timothy
;   REGISTRATION NUMBER:  38,241
;   REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
;   TELEPHONE:  919-541-8587
;   TELEFAX:  919-541-8689
INFORMATION FOR SEQ ID NO:  730:
SEQUENCE CHARACTERISTICS:
;   LENGTH:  635 base pairs
;   TYPE:  nucleic acid
;   STRANDEDNESS:  single
;   TOPOLOGY:  linear
MOLECULE TYPE:  DNA (genomic)
ORIGINAL SOURCE:
;   ORGANISM:  PAG1476UP
US-08-998-416-730
```

Alignment Scores:
Pred. No.:
Score: 2.45e-56
Percent Similarity: 530.50
Best Local Similarity: 62.44%
Query Match: 51.22%
DB: 21.13%
              3

Length: 635
Matches: 105
Conservative: 23
Mismatches: 33
Indels: 44
Gaps: 2

US-10-622-516-2 (1-478) x US-08-998-416-730 (1-635)

```
Qy   29 GlnTyrValAlaAspGlnAlaGlnLeuGlyLysTyrAspGlyValAspAlaGlyLysTyrThr 48
        |||:|
Db   20 CAGTGCGTGAACCAAGAGCATTGGAGGGCGTTATGAGCGGCGTTGCGAAGGCCAAGTACACT 79

Qy   49 IleGlyTyrLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGlyLeuArgSerIleAsnGlnSerLeu 68
        |||||:||||||||
Db   80 ATCGGCTTGGGCCCAGACCAACATGAGCTTTGTGAACGACGCCGGAGGACATCTTACTCGATG 139

Qy   69 CysMetThrValValGluAsnLeuMetGluArgAsnAsnLeuSerTyrAspCysIleGly 88
        |||||||
Db  140 TGTTTGACCGCGTGCTCGAACTTGATGAGAACTACGATATCAGACGGGAAAGCATCGGC 199

Qy   89 ArgLeuGluValGlyThrGluThrThrIleIleAspLysSerErIleSerValLysThrAsnLeu 108
        ||||||:::::||||
Db  200 CGCCTCCGAGGTGGGTACCGGAGACGTTGGCTTCGAAGGTGCAGAGAGTCTATTTTTG 259

Qy  109 MetGlnLeuPheGluGlyLeuSerGlyAsnThrAspIleGluGlyIleAspThrThrAsnAla 128
        |||||||
Db  260 ATGCAGTTGTTCGGCGAG-----AACACCGACTTGGAGGGTGTCGATACCGTGAACGCC 313

Qy  129 CysTyrCysGlyThrAlaAlaValPheAsnAlaValAsnTrpIleGluSerSerSerTrp 148
        |||:||:|
Db  314 TGCTACTGGCGGTACTAACCGGTTGTTAACTCCCTGAACTGGATCGAGTCCAGTTCGTGG 373

Qy  149 AspGly----------- 150
        |||:|
Db  374 GACGGTTCGTGACGCAATCCGTTGTTTGTGGTGACATCCCAAATCACGACAAGGGTGCCCGCC 433

Qy  150 ----------------- 150
Db  434 CGGCCCACTGCGGCGGTGCGGAACGTTCGCCTCTCCGATCGGGTTCCAGACGCGCCCCCCATTGTC 493

Qy  151 -------LeuArgGlyThrHisMetGlnHisAlaThrTyrAspPheTyrLysProAspMet 167
        :::|||
Db  494 TTTGACTCTGTCCGTGCGCTGCTCGTACATGGAGCAGGTCTACGACTTCTACAAGCCTGACTTC 553

Qy  168 LeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSer-AlaIle 187
        :::|||
Db         187
```

Alignment Scores:
Pred. No.:
Score: 1.13e-39
Percent Similarity: 395.00
Best Local Similarity: 85.85%
Query Match: 81.13%
DB: 15.73%
              4

Length: 472
Matches: 86
Conservative: 5
Mismatches: 12
Indels: 4
Gaps: 0

US-10-622-516-2 (1-478) x US-09-833-381-1648 (1-472)

```
Qy  376 ValAlaProAspValPheAlaGluAsnMetLys-LeuArgGluAspThrHisHisLeuVa 395
        |||||||
Db    8 GTGGCACCGGATGTCTTTGCTGAAAACATGAAATGTCAGAGAGGACACACACTTAGC 67

Qy  395 lAsnTyrIleProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValAr 415
        :::||:
Db   68 CAACTATATTCCCCATGT-TCAATANACTCACTCTTTGAAGGAACGTGGTATCTGGTCAG 126

Qy  415 gValAspGluLysHisArgArgThrTyrAlaAargArgProThrProAsnAspAspThrIe 435
        ||||||
Db  127 AGTGGATGAAAAACACAGAAGGAGACTTACGCCGCGGGCCCTTCACAAATGACCACAGTTT 186

Qy  435 uAspGluGlyValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAl 455
        |||||:::
Db  187 GGATGAAGGAAATGGNNCTCGTGCGATAGTAACACAGAGCATAATTCCAAGCCCCTGC 246

Qy  455 aLysLysValProArgLeuProAlaThrAlaAlaGluProGluAla-AlaValIleSer- 474
        ||::||
Db  247 TAAGAAAGTGCCAAGACTCCCTGCCACCCTGCGCGGCCAAATCTGAATCACNCTGCTCATCACAT 306

Qy  475 AsnGlyGluHis 478
        |||||:||
Db  307 AACGGGAGCAC 318
```

```
Db  554 CGCGAGTCGAGTATCCATACGTGGACGGCGCCACTTCTCACTAACATGCTAACGTCAAGGGCCGT 613
        ||||||||||||       |||:::       ||||||||:       ::
Qy  187 uAspArgCysTyr 191
        ||:||:
Db  614 CGACCAGGCTTAC 626
```

NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
  ADDRESSEE: GENOME THERAPEUTICS CORPORATION
  STREET: 100 Beaver Street
  CITY: Waltham
  STATE: Massachusetts
  ZIP: 02354
  COUNTRY: USA
COMPUTER READABLE FORM:
  MEDIUM TYPE: CD/ROM ISO9660
  COMPUTER: PC
  OPERATING SYSTEM: <Unknown>
  SOFTWARE: ASCII
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/107,532A
  FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 60/085,598
  FILING DATE: 14 May 1998
  APPLICATION NUMBER: 60/051571
  FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
  NAME: Ariniello, Pamela Deneke
  REGISTRATION NUMBER: 40,489
  REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (781)893-5007
  TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1715:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1167 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: circular
  MOLECULE TYPE: DNA (genomic)
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
  ORGANISM: Enterococcus faecium
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION 1...1167
SEQUENCE DESCRIPTION: SEQ ID NO: 1715:
US-09-107-532A-1715

Alignment Scores:
Pred. No.:
Score: 5.16e-29
Score: 315.00
Percent Similarity: 40.10%
Percent Identity: 24.82%
Best Local Similarity: 12.54%
Query Match: 4

Length: 1167
Matches: 104
Conservative: 64
Mismatches: 135
Indels: 116
Gaps: 13

US-10-622-516-2 (1-478) x US-09-107-532A-1715 (1-1167)

[Sequence alignment rows omitted — nucleotide/protein alignment data]

RESULT 14
US-08-956-171B-364
; Sequence 364, Application US/0895671B
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;   Gil H. Choi

```
Patrick S. Dillon
Craig A. Rosen
Steven C. Barash
Michael R. Fannon
NUMBER OF SEQUENCES: 5256
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Human Genome Sciences, Inc.
  STREET: 9410 Key West Avenue
  CITY: Rockville
  STATE: Maryland
  COUNTRY: USA
  ZIP: 20850
COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
  COMPUTER: HP Vectra 486/33
  OPERATING SYSTEM: MSDOS version 6.2
  SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/956,171E
  FILING DATE: 20-Oct-1997
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 60/009,861
  FILING DATE: January 5, 1996
  APPLICATION NUMBER: 08/781,986
  FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
  NAME: Mark J. Hyman
  REGISTRATION NUMBER: 46,789
  REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (301) 309-8439
  TELEFAX: (240) 314-1224
INFORMATION FOR SEQ ID NO: 364:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 10813 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
  SEQUENCE DESCRIPTION: SEQ ID NO: 364:
US-08-956-171E-364
```

```
Alignment Scores:
Pred. No.:                           No.:   1.24e-25
Score:                                      301.50
Percent Similarity:                         40.28%
Best Local Similarity:                      24.64%
Query Match:                                12.01%
DB:

Length:        10813
Matches:       104
Conservative:  66
Mismatches:    135
Indels:        117
Gaps:          14
```

```
US-10-622-516-2 (1-478) x US-08-956-171E-364 (1-10813)

Qy   17 ValGlyIleValAlaLeuGlyIleLeuTyrPheProSerGlnTyrValAspGlnAlaGluLeu 36
           ::||||||                      ::|
Db 3831 ATAGGTATCCGATAAAAATAAACTTTACGTTCCAAAGTACTATGGCTGGGTAAAATTA 3890

Qy   37 GlyLeuTyrArgGlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMet 56
                       |||::
Db 3891 GCAGAAGCACGCCAAGTAGACCCAAACAAATTTTAATTGGAATTGGTCAAACTGAAATG 3950

Qy   57 GlyPheCysThrAspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeu 76
                  :||||||::
Db 3951 GCTGTTAGTCCTGTAAACCAAGACATCGTTTCAATGGGCCGCTAAGGCTGCCTAAGGACATT 4010

Qy   77 MetGluArgAsnAsnLeuSerTyrAspCysIleGlyArgLeuLeuValGlyThrGlnThr 96
               :||||||||::
Db 4011 ATAACAGACGAGACAAAAAGAAA------ATTGGTATGGTAATTGTGGCAACTGAATCA 4064

Qy   97 IleIleAspPheLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGly 116
            :|||:::
Db 4065 GCAGTTGATGCTGCTAAAGCAGCCGCGCTGTTCAAATTCACACAACTTATTA--------- 4112

Qy  117 AsnThrAspIleGluGlyIle------------------------------AspThrThrAsnAlaCys 129
                                           ::||||
Db 4113 -------------------------GGTATCAACCTTTGCACGCTGCTTTGAAATGAAGAAGCTTGT 4157

Qy  130 TyrGlyGlyThrAlaAlaValPheAsnAlaValAlaAsnTrpIleGlu----------- 144
        |||                            :::
Db 4158 TATGCTCCAACACCAGCAATTCAATTACGTAAAGATTATTTAGCAACTAGACCGAAATGAA 4217

Qy  145 -------------------------------SerSerSerTrpAspGly 150
                                        |||:::
Db 4218 AAAGTATTAGTTATTCCTACAGATACAGCACACGTTATGGATTGAATTCAGGGGGGAGCCA 4277

Qy  151 LeuArgGly--------------------- 153
        :::|||
Db 4278 ACACAACAAGGTGCTGGGCCCAGTTGCCAGTGTTATTGCACATAATCCAAGCCATTTGGCATTA 4337

Qy  154 ----------ThrHisMetGlnHisAlaTyrAspPheTyrLysProAspMetLeu 168
                  :::                       |||||||
Db 4338 AATGAAGATGCTGTTGCTACACTGAAGACGTTTATGAGTTTATGGTTCGGCGTCCAACTGGA--- 4394

Qy  169 SerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAlaLeuAsp 188
        ::|||:|||                       :::
Db 4395 CATAAAATATTCCATTACAGTTGATGGTGCCGTATTCTAAAGAGTGCTTAATATCCGTCATTCCAA 4454

Qy  189 ArgCysTyrSerValTyrCysLysLysIleHisIleAlaGlnTrpGlnLysGlyLysAsnAsp 208
        :|||      ||||
Db 4455 CAAAGCTGGAATGAAATACGCAAAA-----------------CGTCAAGGTAAG--- 4490

Qy  209 LysAspPheThrLeuAsnAspPheGlyPheMetIlePhePheHisSerSerProTyrCysLysLeu 228
        ||||||||                       :::
Db 4491 ------------TCCGTAGCTGACTTCCCATCTCTATCGTTCCCAGTTCCATTTACACAAAATG 4541

Qy  229 ValGlnIleLysSerLeuAlaArgMetLeuLeuLeuAsnAspPheLeuAsnAspArgAsp 248
                 :::||||||
Db 4542 GGTAAAAAGGCATTAGAGTCAATCATTGATACCGCTGATGAAAACAACTCAAGAGCGT--- 4598

Qy  249 LysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLysGluLeuGluAspThrTyr 268
        |||
Db 4599 ------------TTACGTTCAGGATATGAAGAGATGCTGTAGAT--------------- 4628

Qy  269 PheAspArgAspValGlyLysAlaPheMetLysAlaSerSerGluLeuPheSerGlnLys 288
             :::                          |||
Db 4629 TATAACCGTTATGTC------------------------------------------ 4643

Qy  289 ThrLysAlaSerLeuLeuValSerAsnGlyAsnMetTyrThrSerSerValTyr 308
        ||||                           :::|||
Db 4644 -----------GGTAATATTTAACCTGAGATCATTATAT 4670

Qy  309 GlySerLeuAlaSerValLeuLeuAlaGlnTyrSerProGlnLeuLeuAlaGlyLysLysArgIle 328
        |||                            :::
Db 4671 TTAAGCCTAATATCATTACTTGAAAATCGAGATTTACAA----------GCTGGTGAAACAATC 4724

Qy  329 GlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSerLeuLysValThrGln 348
        ||||||                         :::|||
Db 4725 GGTTTATTCAGTTATGGCTCAGGTTCAGTTGGTGAATTTTATACTAGTTGGACAATTAGTTGAA 4784

Qy  349 AspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeuCysLysLeuSerSer 368
        |||                            :::
Db 4785 GGCTACAAA------------GATCATTTAGATCAAGCTGCACAT----------AAAGCA 4823

Qy  369 ArgLeuAspSerArgThrGlyValAlaProAspValPheAlaGlyAsnMetLysLeuArg 388
        |||                            :::
Db 4824 TTATTAAATAAACCGTACTGAAGTATTCTGTTGATGCATATGAAAAACATTCTTCAAACGTTTT 4883

Qy  389 GluAspAsp 390
             :::|||
Db 4884 GATGAC 4889
```

```
RESULT 15
US-09-134-001C-1458
; Sequence 1458, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
```

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1458

; SEQ ID NO 1458
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1458

Alignment Scores:
Pred. No.:
Score: 4.01e-25
Score: 284.50
Percent Similarity: 38.70%
Percent Identity: 38.70%
Best Local Similarity: 23.26%
Query Match: 11.33%

DB:
Length: 1311
Matches: 107
Conservative: 71
Mismatches: 149
Indels: 133
Gaps: 16

US-10-622-516-2 (1-478) x US-09-134-001C-1458 (1-1311)

```
Qy      16  AspValGlyIleValAlaAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGlu  35
                ::|||||||||:||:|||::|||| ||::|
Db     148  AATATAGGTATAGATAATAAAATAAGTTCTAATGTACCCAAATATTATGTAGACATGGCTAAA  207

Qy      36  LeuGluLysTyrArgAspGlyValAspAlaGlyIleLysTyrThrIleGlyLeuGlyGlnAlaLys  55
                |||| :::                    :::
Db     208  CTTGCAGAAGCGCGCCCAAGTCGATCCTAATAATAATTTTTAATTGGAATTGGTCAAACTGAA  267

Qy      56  MetGlyPheCysThrArgAspArgGluAsnArgLysSerLeuCysMetThrValValGlnAsn  75
                ||||:
Db     268  ATGGCTCGTGAGCCCAGTGAATCAAGATATCGTATCTATGGAGCCAATGCTGCTAAAGAT  327

Qy      76  LeuMet------------GluArgAsnAsnLeuSerTyrArgAspCysIleGlyArgLeuGluVal  92
                :::    |||
Db     328  ATTATAACAGAGAAGAGATAAAAAAGAAT-----------AATTGGTATGGTTATAGTA  372

Qy      93  GlyThrGluThrIleIleIleAspLysSerLysThrAsnLeuMetGlnLeuPhe  112
                :||:||:
Db     373  GCAACTGAGTCTCCGATTGATAATGCCAAAGCAGCAGCCGGTTCAAATTCACCATCTTTTA  432

Qy     113  GluGluSerGlyAsnThrAspIleGluGlyIle-----------AspThr  125
                                                             :::
Db     433  ------------------GGTATTCAACCCTTTGCAAGATGCTTTGAAATG  465

Qy     126  ThrAsnAlaCysTyrGlyThrAlaAlaValPheAsnAlaValAlaAsnTrpIleGluSer  145
                |||||||:                ::|||
Db     466  AAAGAGGCTTGTTATGCAGCAACACCTGCAATTCAACTTGCCAAAGATATCTGCTCAA  525

Qy     146  ---------------------------------------Ser  146
                                                      |||
Db     526  CGCCCTAACGAAAAAGGTTCTTGTCATTGCTGTAGTCACAGCTCGTTATGGTATTCATCT  585

Qy     147  SerTrpAspGlyLeuArgGly---------------------  153
                ::|                          
Db     586  GGTGGTGAGCCTACTCAAGGTGCCGGGTGCGGTGTCGATGTGT  645

Qy     154  -----------ThrHisMetGlnHisAlaThrAspPheTyrLys  164
                            ||||||:
Db     646  ATTTAAAACTTAATGATGCCGTAGCATATACTGAAGACGTTTATGATTTCTGGCGT  705

Qy     165  ProAspMetLeuSerGluThrTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeu  184
                :||||||||||:                    ::|||
Db     706  CCAACGGGT---CATCAATATCCTTAGTTGCTCGGTCATTGTCGAAAGATGCCTATATC  762

Qy     185  SerAlaLeuAspArgCysTyrSerValTyrCysGlyLysLysIleHisAlaGlnTrpGlnLys  204
                |||                              ||:          |||    :::
Db     763  -------------------------------------------AAGTCATTCCAAGAAAGTTGGAATGAA  789
```

```
Qy     205  GluGlyAsnAspLysGlyAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerPro  224
                |||||                |||:          ::|
Db     790  TATGCACGTCGCCATAATAAAACACTCGCTGCTTCACTATGTTTCCATGTACCA  849

Qy     225  TyrCysLysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAsp  244
                :::    |||||||||:::|||||:::|||              :::
Db     850  TTCACCACAAAAATGGGACAAAAAGCTTTAGATTCTATTATAATCATGCCGATGAAACTACA  909

Qy     245  GlnAsnArgAspLysAspLeuAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLeu  264
                ||||||||                  :::            ||:
Db     910  CAAGACCGTCTTAACTCTATTAGTTACCAAGATGCAGTTGAT---------------------  948

Qy     265  GluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGluLeu  284
                :::                 |||||||||||
Db     949  --------------------TATAATCGTTATGTC-------------------------  963

Qy     285  PheSerGlnHisLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyrThr  304
                |||||||||||||||      :::
Db     964  ----------------------------------GGTAATATTTACACA--------  978

Qy     305  SerSerValTyrGlySerLeuAlaSerValLeuLeuAlaGlnTyrSerProGlnGlnLeuAla  324
                ||||||||::                      ||:
Db     979  GGGTCCTTATATTTATTAAGTTCCATCTCTTTATTAGAA-----------ACACGTGATTTAAAAGGC  1032

Qy     325  GlyLysArgIleGlyValPheSerTyrGlySerGlyLeuValLeuAlaAlaThrLeuTyrSerLeu  344
                |||                              ||:::
Db    1033  GGACAAACGATTGGTCTCTTCTTTAGTTATGGTTCTGGTTTCTGGTTGTGATGAAAT  1089

Qy     345  LysValThrGlnAlaThrProGlyLysSerValLeuAlaGluAspTyrIleThrAlaSerLeuCys  364
                |||||||||                              |||:::
Db    1090  -------------------------GGAACAATAGTAGATGGATAGGAGGACAATTA---  1122

Qy     365  AspLeu---------LysSerArgLeuAspSerArgThrGlyValAlaAlaProAspValPhe  381
                |||:|:                  |||:
Db    1123  GATGTTGAGGCGCCCACAAATCTTTATTAATAATACTATAGAGGTTTTCTGTTGATGAAATAT  1182

Qy     382  AlaGluAsnMetLysLeuArgGluAspThrHisSerHisLeuValAsnThrIleProGlnGly  401
                |||                              ||:::
Db    1183  ---GAACATTTTCAAACGCTTTGACCAATTAGAATTCATGAACTTGAAAAATCA  1239

Qy     402  SerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlyLysHisArg  421
                |||:::                    |||
Db    1240  AATGCAGAT-----------CGTGACATTTCTATTAAAATCTATTGATAACAAATATTCGT  1290

Qy     432  
Db    1290  
```

Search completed: June 24, 2004, 16:00:40
Job time : 178 secs

OM protein - protein search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 24, 2004, 13:02:01 ; Search time 49 Seconds
                                            (without alignment)
                                            938.358 Million cell updates/sec

Sequence: 1 MPGSLPLNAEACWPKDVGIV..........PRLPATAEPEAAVISNGEH 478
Perfect score: 2511
Title: US-10-622-516-2

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Maximum DB seq length: 200000000
Minimum DB seq length: 0

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_78:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2480	98.8	520	2	S45497	hydroxymethylgluta
2	2439	97.1	520	2	S21197	hydroxymethylgluta
3	2362	94.1	520	2	S12736	hydroxymethylgluta
4	2352	93.7	520	2	A25332	hydroxymethylgluta
5	2095	83.4	522	2	S13887	hydroxymethylgluta
6	1513	60.3	508	2	S71623	hydroxymethylgluta
7	1494	59.5	508	2	A35865	hydroxymethylgluta
8	1493	59.5	471	2	B55729	hydroxymethylgluta
9	1352.5	53.9	453	2	S38986	hydroxymethylgluta
10	1272.5	50.7	455	2	A53565	hydroxymethylgluta
11	1009.5	40.2	474	2	T10688	hydroxymethylgluta
12	983.5	39.2	454	2	T49718	probable hydroxyme
13	980	39.0	461	2	T09341	hydroxymethylgluta
14	950.5	37.9	491	2	S58202	hydroxymethylgluta
15	900	35.8	447	2	S61875	hydroxymethylgluta
16	705	28.1	462	2	T25726	hypothetical prote
17	338.5	13.5	98	2	S24432	hydroxymethylgluta
18	312.5	12.4	445	2	A84314	3-hydroxy-3-methyl
19	301.5	12.0	388	2	C90059	3-hydroxy-3-methyl
20	299.5	11.9	384	2	D8821	hydroxymethylgluta
21	272.5	10.9	388	2	AE1614	hydroxymethylgl
22	267.5	10.7	388	2	AG1251	hydroxy-3-methylgl
23	244	9.7	398	2	C95201	hydroxymethylgluta
24	241	9.6	398	2	A98068	hypothetical prote
25	211	8.4	412	2	AG0177	probable hydroxyme
26	204	8.1	407	2	B70185	probable hydroxyme
27	158	6.3	42	2	S13490	hydroxymethylgluta
28	147	5.9	345	2	E69678	involved in polyke
29	145.5	5.8	349	2	F69205	3-hydroxy-3-methyl
30	144	5.7	350	2	F71113	probable acyl carr
31	138.5	5.5	350	2	T45082	acyl carrier prote
32	126.5	5.5	350	2	E75047	acyl carrier prote
33	123.5	4.9	716	2	JU0275	beta-galactosidase
34	122	4.9	345	2	A64493	acyl carrier prote
35	114.5	4.6	458	2	C90513	conserved hypothet
36	114	4.5	1486	2	AI0906	glutamate synthase
37	112	4.5	549	2	C89930	alpha-D-1,4-glucos
38	108.5	4.3	1619	2	T18499	hypothetical prote
39	108	4.3	464	2	A87675	armadillo-like pro
40	107.5	4.3	776	2	H71688	carboxyl-terminal
41	106.5	4.2	776	2	T03218	DNA topoisomerase
42	106.5	4.2	1517	2	F55112	glutamate synthase
43	106	4.2	367	2	T20271	hypothetical prote
44	105.5	4.2	1144	2	A54810	TMV resistance pro
45	105.5	4.2	1517	2	C91140	glutamate synthase

ALIGNMENTS

RESULT 1
S45497
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic, adrenal isoform - human
N;Alternate names: 3-hydroxy-3-methylglutaryl coenzyme A synthase
C;Species: Homo sapiens (man)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-May-2000
C;Accession: S45497; I39355
R;Rokosz, L.L.; Boulton, D.A.; Butkiewicz, E.A.; Sanyal, G.; Cueto, M.A.; Lachance, P.A.;
Arch. Biochem. Biophys. 312, 1-13, 1994
A;Title: Human cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase: expression, p
A;Reference number: I39355; MUID:94304197; PMID:7913309
A;Accession: S45497
A;Molecule type: mRNA
A;Residues: 1-520 <ROK>
A;Cross-references: EMBL:L25798; NID:g410027; PIDN:AAA62411.1; PID:g410028
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; oxo-acid-lyase
F;129/Active site: Cys (covalent substrate-binding) #status predicted

Query Match   98.8%; Score 2480; DB 2; Length 520;
Best Local Similarity   91.9%; Pred. No. 4.4e-177;
Matches 478; Conservative   0; Mismatches   0; Indels 42; Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALELYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60

Qy   61 DREDINSLCMTVQONLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVQONLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 180
           ||||||||||||||| ||||||||||||||||||||
Db   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD----------------------------- 149

Qy   150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSYYCKKI 198
                       ||||||||||||||||||||||||||||||||||||||
Db   121 IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSYVCKKI 180

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 240

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 300

Qy   319 SPQOLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
           ||
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

RESULT 2
S27197
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic, fibroblast isoform - human
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-May-2000
C;Accession: S27197; S21590
R;Russ, A.P.; Ruzicka, V.; Maerz, W.; Appelhans, H.; Gross, W.
Biochim. Biophys. Acta 1132, 329-331, 1992
A;Title: Amplification and direct sequencing of a cDNA encoding human cytosolic 3-hydrox
A;Reference number: S27197; MUID:93041939; PMID:1356203
A;Accession: S27197
A;Molecule type: mRNA
A;Residues: 1-520 <RUS>
A;Cross-references: EMBL:X66435; NID:g30008; PIDN:CAA47061.1; PID:g30009
A;Experimental source: fibroblast
C;Function:
A;Description: catalyzes the condensation of acetyl coenzyme A (Ac-CoA) with acetoacetyl
tase
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; oxo-acid-lyase
F;129/Active site: Cys (covalent substrate-binding) #status predicted

Query Match   97.1%;   Score 2439;   DB 2;   Length 520;
Best Local Similarity 90.9%;   Pred. No. 5e-174;
Matches   471;   Conservative   1;   Mismatches   4;   Indels   42;   Gaps   1;

```
Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-----------------------------  149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL  180

Qy  150  --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI    198
                 |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLINDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  HAQWQKEANDNDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLINDQNRDKNSIYSGLKA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  SPQHLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAQ  420

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         ||||||||||||||||||||||||||||||||||||||||
Db  481  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  520
```

RESULT 3
S12736
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), cytosolic - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-May-2000
C;Accession: S12736
R;Ayte, J.; Gil-Gomez, G.; Hegardt, F.G.
Nucleic Acids Res. 18, 3642, 1990
A;Title: Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-meth)
A;Reference number: S12736; MUID:90301491; PMID:1972979
A;Accession: S12736
A;Molecule type: mRNA
A;Residues: 1-520 <AYT>
A;Cross-references: EMBL:X52625; NID:g55946; PIDN:CAA36852.1; PID:g55947
A;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; cytosol; oxo-acid-lyase

Query Match   94.1%;   Score 2362;   DB 2;   Length 520;
Best Local Similarity 87.5%;   Pred. No. 2.8e-168;
Matches   455;   Conservative   11;   Mismatches   12;   Indels   42;   Gaps   1;

```
Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db   61  DREDINSLCITVVQKLMERNNLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-----------------------------  149
         ||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYASGNARPTGGVGAVALL  180

Qy  150  --------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI    198
                 |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IGPNAPVIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYCKKI  240

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLINDQNRDKNSIYSGLEA  258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  RAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLINDQNRDKNSIYSGLEA  300

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  FGDVKLEDTYFDRDVEKAFMKASAELFNQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db  361  SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP  420

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db  421  DVFAENMKLREDTHHLANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHSILDEG  480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         ||||||||||||||||||||||||||||||||||||||||
Db  481  VGLVHSNIATEHIPSPAKKVPRLPATGGEPESAAVISNGEH  520
```

RESULT 4
A25332
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Chinese hamster
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 05-May-2000
C;Accession: A25332
R;Gil, G.; Goldstein, J.L.; Slaughter, C.A.; Brown, M.S.
J. Biol. Chem. 261, 3710-3716, 1986
A;Title: Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the hamster.
A;Reference number: A25332; MUID:8614016; PMID:2869035
A;Accession: A25332
A;Molecule type: mRNA
A;Residues: 1-520 <GIL>

```
A;Cross-references: GB:L00334; NID:g191380; PIDN:AAA37076.1; PID:g387072
A;Note: the source is Chinese hamster ovary cells
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; endoplasmic retic

Query Match   455;   Score 2352;   DB 2;   Length 520;
Best Local Similarity   93.7%;   87.5%;   Pred. No. 1.5e-167;
Matches   455; Conservative   9; Mismatches   14; Indels   42; Gaps   1;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGPCT   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGPCT   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCLTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------      149
          ||||||||||||||||||||||||||||
Db  121 EGIDTTNACYGGTAAVFNAVNWIESSSWDGRYALVVAGDIAIYATGNARPTGGVGAVALL  180

Qy  150 -----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                     ||||||||||||||||||||||||||||||||||||||||||||| |||
Db  181 IGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYRKKI  240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKDNSIYSGLEA 258
          :||||||||| ||||||||||| ||||||||||||||| |||||||||||||||||||||
Db  241 RAQWQKEGNDNDFTLNDFGFMISHSPYCKLVQKSLARMFLNDFLNDQNRDKDNSIYSGLEA 300

Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  301 FGDVKLEDTYFDRDVEKAFMKASSELFNQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  360

Qy  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db  361 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP  420

Qy  379 DVFAENMKLREDTHHLVNYIPQQSIDSLFEGTWYLVRVDEKHRRTYARRPTENDDTLDEG  438
          ||||||||||||||| |||||| ||||||||||||||||||||||||||| ||  | | |
Db  421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPSTNDHNLGDG  480

Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                     478
          |||||| |||||||||||||||||||||   | ||||||
Db  481 VGLVHSNTATEHIPSPAKKVPRLPATAAEBSESAVISNGEH                    520
```

```
RESULT 5
S13887
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - chicken
C;Species: Gallus gallus (chicken)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-May-2000
C;Accession: S13887
R;Kattar-Cooley, P.A.; Wang, H.H.L.; Mende-Mueller, L.M.; Miziorko, H.M.
Arch. Biochem. Biophys. 283, 523-529, 1990
A;Title: Avian liver 3-hydroxy-3-methylglutaryl-CoA synthase: distinct genes encode the
A;Reference number: S13887; MUID:91112772; PMID:1980405
A;Accession: S13887
A;Molecule type: mRNA
A;Residues: 1-522 <KAT>
A;Cross-references: EMBL:M60657
A;Note: the authors translated the codon CTG for residue 280 as Ile
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match   399;   Score 2095;   DB 2;   Length 522;
Best Local Similarity   83.4%;   76.4%;   Pred. No. 2.3e-148;
Matches   399; Conservative   37; Mismatches   42; Indels   44; Gaps   3;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGPCT   60
          ||||||| :  |||||||||||||||||||||| |||||||||||||||||||| ||||:
Db    1 MPGSLPVNTESCWPKDVGIVALEIYFPSQYVDQTELEKYDGVDAGKYTIGLGQGSKMGFCS   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
```

```
RESULT 6
S71623
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) precursor, mitochondrial - human
C;Species: Homo sapiens (man)
A;Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, ch
C;Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 05-May-2000
C;Accession: S71623; A55729; A55729; S55103
R;Mascaro, C.; Buesa, C.; Ortiz, J.A.; Haro, D.; Hegardt, F.G.
Arch. Biochem. Biophys. 317, 385-390, 1995
A;Title: Molecular cloning and tissue expression of human mitochondrial 3-hydroxy-3-meth
A;Reference number: S71623; MUID:95200282; PMID:7893153
A;Accession: S71623
A;Molecule type: mRNA
A;Residues: 1-508 <MAS>
A;Cross-references: EMBL:X83618; NID:g619876; PIDN:CAA58593.1; PID:g619877
A;Experimental source: liver
R;Boukaftane, Y.; Duncan, A.; Wang, S.; Labuda, D.; Robert, M.F.; Sarrazin, J.; Schappert
Genomics 23, 552-559, 1994
A;Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, ch
A;Reference number: A55729; MUID:95154824; PMID:7851882
A;Accession: A55729
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 38-508 <BOU>
A;Cross-references: GB:U12788; GB:U12789
C;Genetics:
A;Gene: GDB:HMGCS2
A;Cross-references: GDB:342071; OMIM:600234
A;Map position: 1p13-1p12
A;Genome: nuclear
C;Function:
A;Description: catalyses the condensation of acetyl-CoA and acetoacetyl-CoA to hydroxymet
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; mitochondrion; oxo-acid-lyase
F;1-37/Domain: transit peptide (mitochondrion) #status predicted <SIG>
F;38-508/Product: hydroxymethylglutaryl-CoA synthase #status predicted <MAT>
F;166/Active site: Cys (covalent substrate-binding) #status predicted

Query Match   282;   Score 1513;   DB 2;   Length 508;
Best Local Similarity   60.3%;   60.4%;   Pred. No. 5.9e-105;
Matches   282; Conservative   69; Mismatches   74; Indels   42; Gaps   1;

Qy    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGPCT   60
          ||||||:    |||||||:|||||||||||||: |||||||||||||||||||:|||:|
Db    1 MPGSLPVNTESCWPKDVGILALEIYFPSQYVDQTELEKYDGVDAGKYTIGLGQSKMGFCS   60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
```

(right-column alignment continuation of RESULT 5)

```
Db   61 DREDINSLCLTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTVLMQLFEESGNTDV  120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD--------------------------      149
Db  121 EGIDTTNACYGGTAALFNAINWIESSSWDGRYALVVAGDIAIYATGNARPTGGVGAVAL  180

Qy  150 -----------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
Db  181 YSPEHLAQGRISEPSYGSGFAATLYSIRVTQDATPGSALDKITASLSDLKARLDSRKCIA  240

Qy  199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKDNSIYSGLEA 258
Db  241 HAQWQKEGTDRGFTLNDFGFMIFHSPYCKLVQKSLARLIINDFLSDQNAETANGVFSGLE  300

Qy  259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY  318
Db  301 AFGDVKLEDTYFDRDVEKAFMKASEELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQ  360

Qy  319 SPQQLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
Db  361 YSPEHLAQRISEPSYGSGFAATLYSIRVTQDATPGSALDKITASLSDLKSRLDSRTGVA  377

Qy  379 DVFAENMKLREDTHHLVNYIPQQSIDSLFEGTWYLVRVDEKHRRTYARRPTENDDTLDEG  438
Db  378 PDVFAENMKIRQETHHLANYIPQCSVEDLFEGTWYLVRVDEKHRRTYARRPVWGDGPLEA  437

Qy  439 VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH                     478
Db  438 GVGLVHSNIATEHIPSPAKKVPRLPATA-AEPEAAVISNGEH                   478

Db  481 GVEVVHPGIVHEHIPSPAKKVPRLPATAESEBGVTVAISNGVH                  522
```

RESULT 7
A35865
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 05-May-2000
C;Accession: A35865; S32477
R;Ayte, J.; Gil-Gomez, G.; Haro, D.; Marrero, P.F.; Hegardt, F.G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3874-3878, 1990
A;Title: Rat mitochondrial and cytosolic 3-hydroxy-3-methylglutaryl-CoA synthases are en
A;Reference number: A35865; MUID:90251660; PMID:1971108
A;Accession: S32477
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-508 <AYT>
A;Cross-references: GB:M33648; NID:g204618; PIDN:AAA41336.1; PID:g204619
R;Gil-Gomez, G.; Ayte, J.; Hegardt, F.G.
Eur. J. Biochem. 213, 773-779, 1993
A;Title: The rat mitochondrial 3-hydroxy-3-methylglutaryl-coenzyme-A-synthase gene conta
A;Reference number: S32477; MUID:93238765; PMID:8097464
A;Accession: S32477
A;Molecule type: DNA
A;Residues: 1-35 <GIL>
A;Cross-references: EMBL:M63800; NID:g294590
A;Note: the sequence shown follows the authors' translation at position 35
C;Genetics:
A;Genome: nuclear
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; cholesterol biosynthesis; coenzyme A; mitochondrion; ox
F;1-37/Domain: transit peptide (mitochondrion) #status predicted <SIG>
F;38-508/Product: hydroxymethylglutaryl-CoA synthase #status predicted <MAT>
F;166/Active site: Cys (covalent substrate-binding) #status predicted

Query Match   59.5%;   Score 1494;   DB 2;   Length 508;
Best Local Similarity   59.1%;   Pred. No. 1.5e-103;
Matches   275;   Conservative   76;   Mismatches   72;   Indels   42;   Gaps   1;

```
Qy       6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDRE   63
             ||||:||||||||| ||||||||||| ||||:|||||||:||||||||||||||||||:
Db      43 PLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI  102

Qy       6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI   65
             ||||:||||||||| ||||||||||| ||||:|||||||:||||||||||||||||||:
Db      43 PLAKTDTWPKDVGILALEVYFPAQVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI  102

Qy      64 DINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGI  123
             |||||||| :|| ||:||||||||||||||||||||||||||||||||||||||||||||
Db     103 DINSLCLTVVQRLMERIQLPWDSVGRLEVGTETIIDKSRAVKTVLMELFQDSGNTDIEGI  162

Qy     101 DINSLCLTVVQRLMERIQLPWDSVGRLEVGTETIIDKSRAVKTVLMELFQDSGNTDIEGI  160
             |||||:::|||:||:||
Db     161 DTTNACYGGTASLFNAANWMESSWDGRYAMVVCGDIAVYPSGNARPTGGAGAVAMLIGP   220

Qy     124 DTTNACYGGTAAVFNAVAVWIESSSWD----------------------------------  149
             ||||||||||| |||| | |:||| ||
Db     161 DTTNACYGGTASLFNAANWMESSWDGRYAMVVCGDIAVYPSGNARPTGGAGAVAMLIGP   220

Qy     150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQ   201
                         |||||||||||||||| |||||||||||||||||||| |||| || ||||:||:
Db     221 KAPLALERGLRGTHMENYYDFYKPNLASEYPIVDGKLSIQCYLRALDRCYTSYRKKIQNQ   280

Qy     202 WQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGD   261
              |::|||||||||||||||||||||||:|||||||||||||||||||||:|||||||||
Db     281 WKQASDRPFTLDDLQYMIFHTPFCKMVQKSLARIMFNDFLSASSDTQTSLYKGLEAFGG   340

Qy     262 VKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQ   321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db     341 LKLEDTYTNKDLDKALLKASQQMFDKKTKASLLYLSTHNGNMYTSSLYGCLASLLSHHSAQ   400

Qy     322 QLAGKRIGVFSYGSGLAATLYSLKVTQDDATPGSALDKITASLCDLKSRLDSRTGVAPDVF   381
             ||| ||||||||||||| :||:||||||||||||||:||||:|||||||||||||||| |
Db     401 ELAGSRIGAFSYGSGLAASFFSFRVSQDDAAPGSPLDKLVSSTSDLPKRLASRKCVSPEEF   460

Qy     382 AENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRP   428
             |:||||||||||||| ||| ||||||||||||||||:|||||||||
Db     461 TEIMNQREQFYHKVNFSPPGDTNSLPPGTWYLERVDEQHRRKYARRP   507
```

RESULT 8
B55729
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), mitochondrial - mouse
C;Species: Mus musculus (house mouse)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 05-May-2000
C;Accession: B55729
R;Boukaftane, Y.; Duncan, A.; Wang, S.; Labuda, D.; Robert, M.F.; Sarrazin, J.; Schappert
Genomics 23, 552-559, 1994
A;Title: Human mitochondrial HMG CoA synthase: liver cDNA and partial genomic cloning, ch
A;Reference number: A55729; MUID:95154824; PMID:7851882
A;Accession: B55729
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-471 <BOU>
A;Cross-references: GB:U12790; GB:U12791
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; mitochondrion; oxo-acid-lyase

Query Match   59.5%;   Score 1493;   DB 2;   Length 471;
Best Local Similarity   59.4%;   Pred. No. 1.6e-103;
Matches   276;   Conservative   74;   Mismatches   73;   Indels   42;   Gaps   1;

```
Qy       6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI   65
             ||||:||||||||| ||||||||||| ||||:|||||||:||||||||||:||||| :||
Db       6 PLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI   65

Qy      66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT  125
             |||| :|| ||:||||||||||||||||||||||||||:||||||||||||||||||||
Db      66 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT  125

Qy     126 TNACYGGTAAVFNAVNVWIESSSWD----------------------------------  149
             ||||||||| ||| | |:||| ||
Db     126 TNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNRPTGGAGAVAMLIGPKA  185

Qy     150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ   203
                         |||||||:||||||:|| |||||||||||||||||||| |||||:  |||||||
Db     186 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKKIQNQWK   245

Qy     204 KEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK   263
             |||||||||||| ||||||||||:|||||||||||||||||||||||||:||||||| ||
Db     246 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLYKGLEAFRGLK  305

Qy     264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQL  323
             |||||||||||||||||||||||||||||||:  ||||||||:|||||||:||||||:
Db     264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQL  323
```

RESULT 9

S38986

hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach
C;Species: Blattella germanica (German cockroach)
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000
C;Accession: S38986
R;Martinez-Gonzalez, J.; Buesa, C.; Piulachs, M.D.; Belles, X.; Hegardt, F.G.
R;Title: 3-Hydroxy-3-methylglutaryl-coenzyme-A synthase from Blattella germanica. Clonin
Eur. J. Biochem. 217, 691-699, 1993
A;Reference number: S38986; MUID:94039108; PMID:7901012
A;Accession: S38986
A;Molecule type: mRNA
A;Status: preliminary
A;Residues: 1-453 <MAR>
A;Cross-references: EMBL:X73679; NID:g416168; PIDN:CAA52032.1; PID:g1772495
A;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match   53.9%; Score 1352.5; DB 2; Length 453;
Best Local Similarity 56.6%; Pred. No. 4.7e-93;
Matches 259; Conservative 66; Mismatches 84; Indels 49; Gaps 4;

```
Qy     13 WPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTV 72
           ||:|||||||||:||||||||| ||||||||| |||| |||||||| :||||||||||||
Db      2 WPSDVGIVALELIFPSQVVDQVDLEVYDNVSAGKYTVGLGQARMGFCTDREDINSLCLTV 61

Qy     73 VQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGG 132
           |||||||||:||||||||||||| ||||||| |||| |||| |||||||| ||||||| |
Db     62 VSRLMERWSIPYSOIGRLEVGTETILDKSKSVKTVLMQLFKD--NTDIEGVDTVNACYGG 119

Qy    133 TAAVFNAVNWIESSSWD----------------------G 150
           |||| |||||||||||||                     |
Db    120 TSALFNAISWVESSSWDGRYALVVAGDIAVYAKGSARPTGGAGAVAMLVGANAPLVFDRG 179

Qy    151 LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVVCKKIHAQWQKEGNDKD 210
           :|||||||||||||||||||||||||||||||||||||||| |||| ||:|||||||:||
Db    180 VRSSHMQHAYDFYKPDLSSLYPVVDGKLSIQCVLSALDHCYQLYCSKI-----QKLGEK- 234

Qy    180 VRSSHMQHAYDFYKPDLSSLYPVVDGKLSIQCVLSALDHCYQLYCSKI-----QKLGEK- 234
```

RESULT 10

A53565

hydroxymethylglutaryl-CoA synthase
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - German cockroach
C;Species: Blattella germanica (German cockroach)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C;Accession: A53565

Query Match   40.2%; Score 1009.5; DB 2; Length 474;

R;Buesa, C.; Martinez-Gonzalez, J.; Casals, N.; Haro, D.; Piulachs, M.D.; Belles, X.; Heg
J. Biol. Chem. 269, 11707-11713, 1994
A;Title: Blattella germanica has two HMG-CoA synthase genes. Both are regulated in the ov
A;Reference number: A53565; MUID:94216267; PMID:7909314
A;Accession: A53565
A;Molecule type: mRNA
A;Status: preliminary
A;Residues: 1-455 <BUE>
A;Cross-references: GB:X75516; NID:g488133; PIDN:CAA54652.1; PID:g488134
A;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match   50.7%; Score 1272.5; DB 2; Length 455;
Best Local Similarity 53.7%; Pred. No. 4.4e-87;
Matches 246; Conservative 64; Mismatches 101; Indels 47; Gaps 4;

```
Qy     11 ACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCM 70
           | ||:|||||||:|||||||||||| ||||||||| :|||||||||||| |||||||||
Db      2 AHWPBDVGIIGIEMIFPSLVVDQAELETYDEVSPGKYTMGLGQDKMGVCTDREDINSLCL 61

Qy     71 TVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACY 130
           |||||||||||| ||:|||||||||||||||||||| |||| ||| |||||||| |||||
Db     62 TAVDKLMERNNIDYNDIGWLEVGTETILDKVKSVKTVLMQLFEESGNTDVEGIDTINACY 121

Qy    131 GGTAAVFNAVNWIESSSSWD----------------------149
           ||||| |||||||||||||
Db    122 RGTAALFNALIWIESSSWDGRYAIVVAADIAIYAKECSPTGGAGALLMLIGANAPIVDR 181

Qy    150 GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVVCKKIHAQWQKEGNDK 209
           |:|||||||||||||||||||||||||||||||||||||||:||||   |:|||||||||
Db    182 GVRASHMKHAYDFYKPDLMSEYPVVDGKLSVQCYLSALDHCYPRFCSKTEKYLKRCGKEN 241

Qy    210 -DFTLNDFGEMIFHSPSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTY 268
           :|||||||||||||||||||||||||||||||||||||||
Db    242 TKIDLDYFDAFVFHSPYCKLVQKSVARLVLNDFI---QYPEK---YQDLQQIRNLKFEDTY 296

Qy    269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQLAGKRI 328
           |||||||||||||||||||||||||||||||||||||||||||| ||||:||||||||
Db    297 FDRDIEKIFMDKSKQLFEBKKTKPSLMLANQVGNMYTPSLYGGLVSLLISEDIGELAGKCI 356

Qy    329 GVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
           |||||||||| |||||||||||||||||||||||||||| |||||||||||||||:||
Db    357 CMFSYGSGFAASMFSLHISTDSSPGSTLSRLVTNLTHIKPQVQQRVRLSPGEFENIMBIR 416

Qy    389 EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYAR 426
           |||||||||||||||||||||||||||||||||||
Db    417 EQNHHKAPYTPVASPNTLFPGTWYLESIDSMHRRXYKR 454
```

RESULT 11

T09688

hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5), ozone-inducible - Scotch pine
A;Alternate names: 3-hydroxy-3-methylglutaryl-coenzyme A synthase
C;Species: Pinus sylvestris (Scotch pine)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T09688
R;Wegener, A.; Gimbel, W.; Werner, T.; Hani, J.; Ernst, D.; Sandermann, H.
Biochim. Biophys. Acta 1350, 247-252, 1997
A;Title: Molecular cloning of ozone-inducible protein from Pinus sylvertris L. with high
A;Reference number: Z16823; MUID:97214637; PMID:9061017
A;Accession: T09688
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-474 <WEG>
A;Cross-references: EMBL:X96386; NID:g1655678; PIDN:CAA65250.1; PID:g1655679
A;Experimental source: tissue-type needles
C;Function:
C;Description: catalyzes the condensation of acetyl coenzyme A (Ac-CoA) with acetoacetyl
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase
C;Superfamily: hydroxymethylglutaryl-CoA synthase
tase

Query Match

```
Best Local Similarity   43.5%; Pred. No. 1.9e-67;
Matches 208; Conservative 76; Mismatches 141; Indels 53; Gaps 7;

Qy    14 PKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV 73
          ||:||:||||:|||:|||:|||||||||:|||||||||||||:||||||||||||||||
Db     5 PENVGILAMEIYFPTTCVQQEDLETFDGVSKGKYTIGLGQDCMTFCTDLEDVISMSLTAV 64

Qy    74 QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDLIEGIDTTNACYGGT 133
          :||: ||:|:|:::||||||:||||||:||||||||:|||||||||:||||||||||||
Db    65 TSLLEKYEIDPKQIGRLEVGSETVIDKSKSIKTWLMHIPEKCGNTEIEGVDSTNACYGGT 124

Qy   134 AAVFNAVANWIESSSWDG---------------------------------------L 151
          |||||| |||||||||||                                       |
Db   125 AALFNCINWIESSSWDGRYGLVAIDSAVYAEGAARPTGGAAAVAMLIGPNAPIATESKY 184

Qy   152 RGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEIGNDKDF 211
          ||||| |||||||:||:|||||||||:||||||||||||:|:||||||||:||||||||:
Db   185 RGTHMAHVYDFYKPNLASEYPVVDGKLSQTCYLMALDSCYKRPCNKFE---KEG--RQF 239

Qy   212 TLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEARFGDVKLEDTYFDR 271
          ||:||||| ||||||||||||:||||| :||:| :|| :||:||| :|||:|:|||:|:
Db   240 SLLDTDYIAFHSPYNKLVQKSFPGRLLFNDFSRHARSVGKDAQEKLEPFAGLSEQDSYNSR 299

Qy   272 DVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVF 331
          |||||||:||||||||||||||::||:|||||||||||||||||||||::|||||:|:|
Db   300 DLEKVSQQLAKPLYDAKIQPSTLFPKQVGNMYTASLYAALASII-HNKHTLLDGQRVNMF 358

Qy   332 SYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDT 391
          |||||||:||:||||:|:|||||||||||||:||:||||||||:||||||||:|:|:|||
Db   359 SYGSGLASTLFSFKIREGQFP-FTLSNIT-EVMDVQNKLDSRHEFLPEDFVENLKRMETL 416

Qy   392 HHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRFFPNDDTLDEGV---GLVHSNI 446
          :: ||: ||||:||||||||||||:|:||::|:|:::|| ||||||:    ||:||:|
Db   417 YGAKDFVSTSQLSLLRPGAFYLTKVDSMYRRFYSRKVISAGDNFEKSKLANGTTHDEL 474


RESULT 12

T49718
probable hydroxymethylglutaryl-CoA synthase [imported] - Neurospora crassa
N;Alternate names: protein B23L21.310
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000
C;Accession: T49718
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Hartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, May 2000
R;Reference number: Z25022
A;Reference number: T49718
A;Molecule type: DNA
A;Status: preliminary
A;Accession: T49718
A;Molecule type: DNA
A;Residues: 1-454 <SCH>
A;Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.310; NCSP:B23L21.310
A;Experimental source: BAC clone B23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.310
A;Genetics:
A;Map position: 6
A;Introns: 20/3; 55/1; 409/3
A;Superfamily: hydroxymethylglutaryl-CoA synthase

Query Match    39.2%;  Score 983.5; DB 2;  Length 454;
Best Local Similarity  44.6%;  Pred. No. 1.6e-65;
Matches 207; Conservative 73; Mismatches 115; Indels 69; Gaps 9;

Qy    14 PKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV 73
          |:|:|||||:|||:||||||||||:|||:|||||||||||||||||||||||||:||||
Db     5 PQNIGIKAIEIYFPSQYVEQSELEKFDGVSTGKYTIGLGQTKMAFCDDREDIYSLAITAV 64

Qy    74 QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDLIEGIDTTNACYGGT 133
          ::|:|:||||||||:|||||||||||||||||||:||||||::|||:|:||||||||||
Db    65 SRLLKNYEIDTNTIGRLEVGTETLLDKSKSVKSVLMQLFGE--NTNIEGVDTINACYGGT 122

Qy   134 AAVFNAVANWIESSSWD------------------------------------------L 151
          |||||||:||||||||:                                         |
Db   123 NAFFNSVNWIESSAWDGRHDAIVVAGDIALYAKGNARPTGGAGCVAMLVGPNAPIAVEPGL 182
```

```
Qy   152 RGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQK------E 205
          ||||| |||||||:||:|||||||||:||||||||||||:|:||||||||:||      |
Db   183 RGYTMAHAYDFYKPDLTSEYPYVDGHYSVSNCTTEALDGAYRAY-----NOREKLLTNGVN 237

Qy   206 GNDKDFT----LNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLL 261
          ||||||:    |||||||||||||||||||||||||:| :||:|:  :|||:::::
Db   238 GHSEDSTKTFDLRFDYLAFHAPFCKLVQKSYARLLLHDYLAN----PESSVFADVPPEVRD 294

Qy   262 VKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQ 321
          |||||||||| |||||||||||||||||||:||||||||||||||||||||||||||||
Db   295 MDYKKSLTDKVVEKIFMTLLTKKRFQERVNPALQVPFTLCGNMYCGSVWGGLASIIGHVDSA 354

Qy   322 QLAGKRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVF 381
          ||||||||:|||||||:||:||||||||||||||||||||:|||:|||||||||:||||
Db   355 QLEGKRIGLFSYGSGLAASFCSPRVT-------GSTEKLAKTL-NLPARLAARRAVPEPSY 407

Qy   382 AENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYA 425
          |||||||||:::||:||||||:|||||||||||||||:||||||
Db   408 DAMCDLRKQAHLQKNYTPKGEVSTLEPGTYYILENVDDMFKRTYS 451


RESULT 13

T09341
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - Arabidopsis thaliana
N;Alternate names: 3-hydroxy-3-methylglutaryl-coenzyme A synthase; protein T26M18
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Jun-2000
C;Accession: T09341; JC4567
A;Molecule type: mRNA
A;Residues: 1-305, S', 307-341,'N', 343-461 <MON>
A;Cross-references: EMBL:X83882; NID:gi1433389; PIDN:CAA58763.1; PID:gi1433390
A;Accession: JC4567
R;Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16650
A;Comment: This enzyme mediates the conversion of three acetyl-CoA molecules to one mole
ivity, and uptakes amino acids.
C;Genetics:
A;Gene: ATSP:T26M18
A;Genetics:
A;Gene: ATSP:T26M18.30
A;Molecule type: DNA
A;Residues: 1-461 <BEV>
A;Cross-references: EMBL:AL078606
A;Experimental source: cultivar Columbia; BAC clone T26M18
R;Montanat, F.; Guilloton, M.; Karst, F.; Delrot, S.
Gene 167, 197-201, 1995
A;Title: Isolation and characterization of a cDNA encoding Arabidopsis thaliana 3-hydrox
A;Reference number: JC4567; MUID:96144274; PMID:8566777
A;Accession: JC4567
A;Molecule type: mRNA
A;Residues: 1-305, S', 307-341,'N', 343-461 <MON>
A;Cross-references: EMBL:X83882; NID:gi1433389; PIDN:CAA58763.1; PID:gi1433390
A;Comment: This enzyme mediates the conversion of three acetyl-CoA molecules to one mole
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; glycoprotein; oxo-acid-lyase
F;117/Active site: Cys (covalent substrate-binding) #status predicted
F;269/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match    39.0%;  Score 980; DB 2;  Length 461;
Best Local Similarity  43.3%;  Pred. No. 2.9e-65;
Matches 198; Conservative 79; Mismatches 128; Indels 52; Gaps 6;

Qy    15 KDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
          |:||||||:||||||||||||||:|||||||||||||:|||||:|| :|||||||||||
Db     3 KNVGILAMDIYFPTCVQQEALEAHDGASKGKYTIGLGQDCLAFCTELEDVISMSFNAVT 62

Qy    75 NLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDLIEGIDTTNACYGGTA 134
          ::|:||||:||||||||||||||||||||||||:||||||:|||||:||||||||||||
Db    63 SLFEKYKIDPNQIGRLEVGSETVIDKSKSIKTFLMQLFEKCGNTDVEGVDSTNACYGGTA 122

Qy   135 AVFNAVANWIESSSWDG---------------------------------------LR 152
          |||||||||||||||||                                       ||
Db   123 ALINCVNWWESNSWDGRYGLVICTDSAVYAEGPARPTGGAAAIAMLIGPDAPIVEBSKLR 182

Qy   153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEIGNDKDFT 212
          ||||||||||||||:|||||:||||||||||||||||:|||||||||||||| :|||||:
Db   183 ... (alignment continues)
```

```
Db  183 ASHMAHVDFYKPNLASEYPVVDGKLSQTCYLMALDSCYKHLCHKNKFE---KIEG--KEFS 237
        ||| |   ||||| |||||| |||||  || |||| | || | || ||    ||     ||
Qy  213 LNDFGFMIFHSPYCKLVQKSLLARMLLNDFLNDQNRDKNSIYSGLEBAFGDVKLEDTYFDRD 272

Db  238 INDADYIVFHSPVNKLVQKSFARLLYNDFLRNASSIDERAKEKFTPYSSITLDESYQSRD 297
        :.::  ::||:| |||||| |||||:|||                    :  |||: :||
Qy  273 VEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVFS 332

Db  298 LEKVSQOIAKPFYDAKVQPTTLIPKEVGNMYTASLYAAFASLIHK-KHNDLAGKRVVMFS 356
        ::| : :| : ||||    || | |||||||:| :| ::||| ||   |:||||:  |||
Qy  333 YGSGLAATILYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDTH 392

Db  357 YGSGSTATMFSLRLNDNDKPFSISN--IASVMDVGGKLKARHEVAPEKFVETMKLMEHRY 414
        ||||  || :||::  :||||:| |    ||:| ||:| || ||| ::| |||||:| ||
Qy  393 HLVNYI--PQGSIDSLFEGTWYLVRVDEKHRRTYARR 427

Db  415 GAKDFVTTKEGIIDLLAPGTYYLKEVDSLIYRFYGKK 451
        |   |:   |||:| || ||||| .||| |||||||
```

RESULT 14
S58202
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YM4987.09c; protein YML126c
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 05-May-2000
C;Accession: S58202
R;Bowman, S.
submitted to the EMBL Data Library, July 1995
A;Reference number: S58194
A;Accession: S58202
A;Molecule type: DNA
A;Residues: 1-491 <BOW>
A;Experimental source: strain AB972
A;Cross-references: EMBL:Z50178; NID:g927528; PIDN:CAA90557.1; PID:g927536; MIPS:YML126c
A;Genetics:
A;Gene: SGD:HMGS
A;Cross-references: SGD:S0004595; MIPS:YML126c
A;Map position: 13L
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

```
Query Match            37.9%;  Score 950.5;  DB 2;  Length 491;
Best Local Similarity  43.4%;
Matches 198;  Conservative 72;  Mismatches 127;  Indels 59;  Gaps 7;

Qy   14 PKDGIVALEIYPSQYVDQAAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV 73
         |.: | :|:|| ||||||  ||||:||| .|||||||||. | | .|||||| ||:|:
Db   46 PQNVGIKGIQIYIPTQCVVNQSELEKFDGVSQGKYTIGLGQTNMSFVNDREDIYSMSLTVL 105

Qy   74 QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGT 133
          .| . :| | . |||||||||:||||||||:.|||||  . ||:||||| ||||||
Db  106 SKLIKSYNIDTNKIGRLEVGTETLIDKSKSVKSVLMQLFGE--NTDVEGIDTLNACYGGT 163

Qy  134 AAVFNAVNWIESSSWDG--------------------------------LR 152
         .. ||::.||||:.||                                 ||
Db  164 NALFNSLMWIESNAWDGRDAIVVCGDIAIYDKGAARPTGCGAGTVAMWIGPDAPIVFDSVR 223

Qy  153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQW-----QKEGNDK 209
         .: |.||||||| :|||| ||||.:|: || .||| . |.|.|||.|       ...||.
Db  224 ASYMEHAYDFYKPDFTSEYPYVDGHFSLTCYVKALDQVVKSYSKKAISKGLVSDDPAGSDA 283

Qy  210 DFTLNDFGFMIFHSPYCKLVQKSLLARMLLNDFLNDQNRDKNSIYSGLEA-FGDVKLEDTY 268
          : |: | .:|| ||||| ||| ||||:|::|                  |||:.. .|
Db  284 LNVLKYFDYNVFHVPTCKLVTKSYGRLLYNDF-------RANPQLFPEVDAELATRDYDESL 338

Qy  269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRI 328
          | ::||.|  |:  ||:| :|.||:| ||:|||||:| | ::|||| .  ::|.|||:
Db  339 TDKNIEKTFVNVAKPFHKERVAQSLIVPFNTGNMYTASVTAAFASLLNTVUGSDDLQGKRV 398

Qy  329 GVFSYGSGLAATILYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
         |.|||||||| :|||| || :|    .| ::|  |  | |||||||.||||||||||||:
Db  399 GLFSYGSGLAASLYSCKIVGD--------VQHIIKEL-DIINLIAKRITETPKDYEAAIELR 451
```

```
Qy  389 EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTY 424
        |:.|||||||||||||.|||||||||:.|||||||:
Db  452 ENAHLKKNFKPQGSIEHLQGSGVYTYLTNIDDKFRRSY 487
```

RESULT 15
S61875
hydroxymethylglutaryl-CoA synthase (EC 4.1.3.5) - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 05-May-2000
A;Accession: S61875; T38843
R;Katayama, S.; Adachi, T.; Takao, K.; Nakagawa, T.; Matsuda, H.; Kawamukai, M.
A;Title: Molecular cloning and sequencing of the hcs gene, which encodes 3-hydroxy-3-met[hyl...]
Yeast 11, 1533-1537, 1995
A;Reference number: S61875; MUID:96353436; PMID:8750242
A;Accession: S61875
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <KAT>
A;Cross-references: EMBL:U32187; NID:g974430; PIDN:AAB17601.1; PID:g974431
A;Note: the authors translated the codon AGT for residue 236 as Arg, TCT for residue 358
R;Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A;Reference number: 221815
A;Accession: T38843
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-447 <GEN>
A;Cross-references: EMBL:Z98530; PIDN:CAB11060.1; GSPDB:GN000066; SPDB:SPAC4F8.14c
A;Experimental source: strain 972h-; cosmid c4F8
A;Genetics:
A;Gene: SPDB:SPAC4F8.14c
A;Map position: 1
A;Introns: 20/3
C;Superfamily: hydroxymethylglutaryl-CoA synthase
C;Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

```
Query Match            35.8%;  Score 900;  DB 2;  Length 447;
Best Local Similarity  41.5%;
Matches 191;  Conservative 72;  Mismatches 133;  Indels 64;  Gaps 6;

Qy   15 KDVGIVALEIYPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
         || |: .|.|| |||.|:| || .||| ..|||||||  : .| .||||| || :.:
Db    6 KDIGIRGLVIYLTPNQVVEQAALEAHDGVSTCKYTIGLGLTTQMAFVDDREDIYSFGLTALS 65

Qy   75 NLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGTA 134
          |::.. :. . ||||||||||||||||||||:.|||||   :. :||||  |||||.
Db   66 QLIKRYQIDISKIGRLEVGTETIIDKSKSVKSVLMQLFGD--NHNVEGIDCVNACYGGVN 123

Qy  135 AVFNAVNWIESSSWD------------------------------------GLR 152
         |.||:.:||||:.||                                    ||
Db  124 ALFNTIDWIESSAWDGRDGIVVAGDIALYAKGNARPTGGAGCVALLVGPNAPIVFEPGLR 183

Qy  153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFT 212
         ||:|||||||||| :|||| ||||.:|: || .||| .|.|...: :.:    .|.:.|.
Db  184 GTYMQHAYDFYKPDLLTSEYPVVDGHFSLECYVKALDGAYANYNVRDVA----KNGKSQGLG 240

Qy  213 LNDFGFMIFHSPYCKLVQKSLLARMLLNDFL-------NDQNRDKNSIYSGLEAFGDVKLEDTY 268
         |: | .:||:| |||.|||: || :|...|        .| . .. :.. : |  .. |
Db  241 LDRFDYCIFHAPTCKQVQKAYARLLYTDSAABPSNPELEGVRELLSTLDA-------KKSL 294

Qy  269 FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRI 328
          | .:|| .  | : |...: .. | :. .|.|||:: | | ::: ||.   ::|.|||
Db  295 TDKALEKGLMAITKERFNKRVSPSVYAPTNCGNMVTASIFSCLTALLSRVPADELKGKRV 354

Qy  329 GVFSYGSGLAATILYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR 388
         | .|||||||| : :| :| :|    .|..| | |   ...: :. |  |  .:|.|:||
Db  355 GAYSYGSGLAASFFSFRVVKGDVSE------IAKKTNLVNDLDNRHCLTPTQYERAIELR 407

Qy  389 EDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRP 428
         ..||||.|:.|.|||:.| .||||||..||: .||||:.|
Db  408 HQAHLKKNFTPKGSIERLRSGTYYLTGIDDMFRRSYSVKP 447
```

Search completed: June 24, 2004, 13:21:29
Job time : 52 secs

```
OM protein - protein search, using sw model

              Copyright (c) 1993 - 2004 Compugen Ltd.

                            GenCore version 5.1.6

Run on:           June 24, 2004, 11:17:16 ; Search time 84 Seconds
                                              1795.449 Million cell updates/sec
Sequence:         1 MPGSLPLNAEACWPKDVGIV..........PRLPATAAEPEAAVISNGEH 478
Perfect score:    2511
Title:            US-10-622-516-2

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:  1017041

Post-processing: Minimum Match 0%
Maximum DB seq length: 0
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :        SPTREMBL_25:*
                  1:  sp_archea:*
                  2:  sp_bacteria:*
                  3:  sp_fungi:*
                  4:  sp_human:*
                  5:  sp_invertebrate:*
                  6:  sp_mammal:*
                  7:  sp_mhc:*
                  8:  sp_organelle:*
                  9:  sp_phage:*
                  10: sp_plant:*
                  11: sp_rodent:*
                  12: sp_virus:*
                  13: sp_vertebrate:*
                  14: sp_unclassified:*
                  15: sp_rvirus:*
                  16: sp_bacteriap:*
                  17: sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result                Query
No.      Score   Match Length  DB   ID       Description
------   -----   ------------  --   --       -----------
1        2403.5   95.7%   509   4   Q8N995   homo sapien
2        2348     93.5%   520   11  Q8JZK9   mus musculu
3        2344     93.3%   520   11  Q8KOI5   mus musculu
4        203      79.8%   520   13  Q7ZYN7   xenopus lae
5        1923.5   76.6%   508   13  Q7ZWE2   brachydanio
6        1494     59.5%   508   4   Q8N7N8   homo sapien
7        1493     59.5%   508   11  Q9DBM4   mus musculu
8        1491     59.4%   508   11  Q9DBK1   mus musculu
9        1337     53.2%   465   5   Q9V7N8   drosophila
10       1164     46.4%   457   5   Q9NDA8   dendroctonu
11       1047     41.7%   463   10  Q8H051   oryza sativ
12       1044     41.6%   450   3   Q9HEI9   phycomyces
13       1029.5   41.0%   268   11  Q8C5F4   mus musculu
14       1013     40.3%   464   10  Q944F8   hevea brasi
15       1009.5   40.2%   474   10  P93773   pinus sylve
16       1009     40.2%   464   10  Q94ET0   hevea brasi
```

```
17       992      39.5%   461   10  Q9MGU3   brassica ju
18       983.5    39.2%   454   3   Q9P5J8   neurospora
19       978      38.9%   468   5   Q86HL5   dictyosteli
20       977.5    38.9%   461   10  Q9FVG2   brassica ju
21       975.5    38.8%   461   10  Q9FVG1   brassica ju
22       973.5    38.8%   453   10  Q9FVG0   brassica ju
23       823      32.8%   406   10  Q8L721   arabidopsis
24       445.5    17.7%   382   5   Q8SR61   encephalito
25       381      15.2%   113   10  Q9M6A5   aerides jap
26       314      12.5%   384   2   Q9FD66   enterococcu
27       312.5    12.4%   445   10  Q9HPJ0   halobacteri
28       306      12.2%   445   17  Q92B67   staphylococ
29       301.5    12.0%   461   16  Q99R90   staphylococ
30       301      12.0%   388   2   Q9FD82   staphylococ
31       299.5    11.9%   384   2   Q9CFA9   lactococcus
32       297.5    11.8%   388   2   Q9FD87   staphylococ
33       288.5    11.5%   388   2   Q9FD76   staphylococ
34       288.5    11.5%   388   16  Q8CN06   staphylococ
35       283      11.3%   388   16  Q9FD71   enterococco
36       282      11.2%   383   16  Q83SL4   enterococcu
37       281.5    11.2%   383   16  Q8EP69   oceanobacil
38       272.5    10.9%   388   16  Q92BU0   listeria in
39       272      10.8%   389   2   Q9KWG1   streptomyce
40       267.5    10.7%   388   2   Q8Y781   listeria mo
41       254      10.1%   388   2   Q8L113   paracoccus
42       244      9.7%    398   16  Q97PB3   streptococc
43       243      9.7%    398   2   Q9FD56   streptococc
44       241      9.6%    398   16  Q8DNS4   streptococc
45       240      9.6%    389   16  Q88V14   lactobacill
```

ALIGNMENTS

RESULT 1
Q8N995

```
ID   Q8N995            PRELIMINARY;      PRT;     509 AA.
AC   Q8N995;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLj38173.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK095492; BAC04559.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
KW   Hypothetical protein.
SQ   SEQUENCE  509 AA;  56238 MW;  75AAEB049EAB1460 CRC64;
```

```
Query Match        95.7%;    Score 2403.5;   DB 4;   Length 509;
Best Local Similarity   89.8%;    Pred. No. 2.2e-168;
Matches   467;  Conservative    0;  Mismatches   0;  Indels   53;  Gaps   2;

Qy     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
```

```
Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-------------------------------  149
         ||||||                                                       
Db  121  EGIDTT--------------NAVNWIESSSWDGRYALVVAGDIAVYATGNARPTGGVGAVALL 169

Qy  150  ---------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                        ||||||||||||||||||||||||||||||||||||||||||||
Db  170  IGPNAPLIFERGLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  229

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  230  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  289

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSYVGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  290  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSYVGSLASVLAQY  349

Qy  319  SPQQLAGKRIGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  350  SPQQLAGKRIGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  409

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  410  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  469

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         ||||||||||||||||||||||||||||||||||||||||
Db  470  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  509
```

RESULT 2
Q8JZK9

```
ID   Q8JZK9;      PRELIMINARY;      PRT;     520 AA.
AC   Q8JZK9;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to 3-hydroxy-3-methylglutaryl-coenzyme A synthase 1
DE   (Hypothetical protein) (Pre B-cell leukemia transcription factor
DE   1).
GN   B130032C06RIK OR HMGCS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium.;
RA   The RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Retina;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC029693; AAH29693.1; -.
DR   EMBL; BC034317; AAH34317.1; -.
DR   EMBL; BC023851; AAH23851.1; -.
DR   EMBL; AK031297; BAC27338.1; -.
DR   EMBL; AK044835; BAC32112.1; -.
DR   EMBL; AK045094; BAC32218.1; -.
DR   MGD; MGI:107592; Hmgcs1.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR0005590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   520 AA;  57569 MW;  E28F8772CF64D85C CRC64;
```

```
Query Match             93.5%;  Score 2348;  DB 11;  Length 520;
Best Local Similarity   86.9%;  Pred. No. 2.8e-164;
Matches  452;  Conservative   13;  Mismatches   13;  Indels   42;  Gaps    1;
```

```
Qy    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db    1  MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT   60

Qy   61  DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120
         |||||||||||||||||||  ::|||||||||||||||||||||||||||||||||||||
Db   61  DREDINSLCMTVVQNLMERHSLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI  120

Qy  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-------------------------------  149
         |||||||||||||||||||||||||||||
Db  121  EGIDTTNACYGGTAAVFNAVNWIESSSWD-------------------------------  149

Qy  150  ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI  198
                      ||||||||||||||||||||||:|||||||||||||||||||||:|||
Db  150  ------------GLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYRKKI  198

Qy  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  199  HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA  258

Qy  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSYVGSLASVLAQY  318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  259  FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSYVGSLASVLAQY  318

Qy  319  SPQQLAGKRIGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  319  SPQQLAGKRVGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||:|||||||||||||||||||||||||||||||||:|||||||||
Db  379  DVFAENMKLREDTHHLIANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPFTNDHSLDEG  438

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         |||||||||||||||||||||||||||:|||:||||||||
Db  439  VGLVHSNIATEHIPSPAKKVPRLPATSAESESSAVISNGEH  478
```

RESULT 3
Q8K0I5

```
ID   Q8K0I5;      PRELIMINARY;      PRT;     520 AA.
AC   Q8K0I5;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   HMGCS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RA   Strausberg R.;
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC031363; AAH31363.1; -.
DR   MGD; MGI:107592; Hmgcs1.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_CoA_SYNTHASE; 1.
```

```
Qy  319  SPQQLAGKRIGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  378
         |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  SPQQLAGKRVGVFSYGSGLAATIYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP  420

Qy  379  DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG  438
         ||||||||||||||||:|||||||||||||||||||||||||||||||||:|||||||||
Db  421  DVFAENMKLREDTHHLIANYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPFTNDHSLDEG  480

Qy  439  VGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH  478
         :|||||||||||||||||||||||||||:|||:||||||||
Db  481  MGLVHSNIATEHIPSPAKKVPRLPATSAESESSAVISNGEH  520
```

KW   Hypothetical protein.
SQ   SEQUENCE   520 AA;   57552 MW;   10F13278A18182262 CRC64;

Qy   Best Local Similarity   93.3%;   Pred. No. 5.5e-164;
Qy   Best Local Similarity   86.9%;   Score 2344;   DB 11;   Length 520;
     Matches 452;   Conservative 12;   Mismatches 14;   Indels 42;   Gaps   1;

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DREDINSLCPTVVQKLMERHSLSYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------- 149
           ||||||||||||||||||:||||||||||:
Db   121 EGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDIAIYATGNARPTGGVGAVALL 180

Qy   150 -------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVYRKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEA 258
           ||||||||||||||||||||||||||||||||||||||||||||||:||||||::|||||
Db   241 RAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLARMFLNDFLNDQNRDKNSIYSGLEA 300

Qy   259 FGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 318
           ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||:
Db   301 FGDVKLEDTYFDRDVEKAFMKASSELFNQKTKASLLVSNQNGNMYTSSVYGSLASVLAQY 360

Qy   319 SPQQLAGRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAP 378
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 SPQQLAGRVGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTCVAP 420

Qy   379 DVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEG 438
           ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 DVFAENMKLREDTHHLANYIPQCSIDSLFEGTWYLVRVDEKHRRTYARRPFTNDHSLDEG 480

Qy   439 VGLVHSNIATEHIPSPAKKVPRLPATSABSESAVISNGEH 478
           ||||||||||||||||||||||||||||||| |||| |||
Db   481 MGLVHSNTATEHIPSPAKKVPRLPATSABSESAVISNGEH 520

Db   Best Local Similarity   73.5%;   Pred. No. 6.9e-139;
Db   Best Local Similarity   41;   Mismatches 53;   Indels 44;   Gaps   3;
     Matches 383;   Conservative 41;

RESULT 4
ID   Q7ZYN7
AC   Q7ZYN7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC042229; AAH42229.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   520 AA;   57227 MW;   4BF61D37CC5DB97A CRC64;

Qy   Query Match   79.8%;   Score 2003;   DB 13;   Length 520;

Qy   Best Local Similarity   76.6%;   Pred. No. 4.7e-133;
     Matches 383;   Conservative 47;   Mismatches 52;   Indels 45;   Gaps   3;

Qy   13 WPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTV 72

ID   Q7ZWB2
AC   Q7ZWB2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Body;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049456; AAH49456.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   508 AA;   55936 MW;   BE7DFB437CF866C6 CRC64;

RESULT 5
Q7ZWB2   PRELIMINARY;   PRT;   508 AA.

Qy   1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCS 60
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||:
Db   1 MPGSLPPNGESSWPKDVGIVALEIYFPSQYVDQEELEKFDGVSAGKYTIGLGQCKKMGFCS 60

Qy   61 DREDINSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
           ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||:
Db   61 DREDINSLCITVVQRLMERNNLSYDCIGRLEVGTETIIDKSKSVKTVLMQLFEESGNTDV 120

Qy   121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------------------------- 149
           |||||||||||||||||:|||||:||||:
Db   121 EGIDTTNACYGGTAALFNAVNWVESSSWDGRYALVVAGDIAVYATGSARPTGGAGAVAML 180

Qy   150 -------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                   ||||||||||||||||||||||:||||||||||||||||||||||||:
Db   181 VGPNATLVCERGLRGTHMQHAYDFYKPDMSESYPVVDGKLSIQCYLSALDRCYSVYRKKI 240

Qy   199 HAQWQKEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNS-IYSGLE 257
           |||:||||||||||||||||||||||||||||||:||||||||||||:|||| |||||||
Db   241 HAQWQKEGKDKDFTLNDFGFMTFHSPYCKLVQKSLARLLVNDFISDSNPNTEGGVYVGLD 300

Qy   258 AFGDVKLEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQ 317
           :||:|||||||||||||||||||||:|||||||||||:||||:|||:||||||||||||:
Db   301 SFRDLKLEDTYFDRDVEKAFLKASTEIFNEKTKASLLVSKENGNMWTPSVYGCLASVLAQ 360

Qy   318 YSPQQLAGRIGVFSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVA 377
           |||||||| |||||||||||||| |||:||||||||||||||||||||||||:||||||
Db   361 YSPQQLAGQRGIGVFSYGSGPAATLYSLRVSQDAMPGSSLDKLTSSLSDLKARLDSRKNVS 420

Qy   378 PDVFAENMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDE 437
           |:|||:||||||||||||||||||||||||||||||||:||||||:|||:|||||| |||
Db   421 PSIFADNMKLREDTHHLANYIPQGSVDDLFPGTWVLVRDEKHRRFYARSSLMSDGPLDA 480

Qy   438 GVGLVHSNIATEHIPSPAKKVPRLPATAAEPEAAVISNGEH 478
           |:|||||||||||||||||||||||||: ||||||||||||
Db   481 APESVLASTANEHFPSPAKKVPRID-PAAAEPISVTNGEH 520

Qy   Query Match   71.7%;   Score 1923.5;   DB 13;   Length 508;
     Matches 365;   Conservative 47;   Mismatches 52;   Indels 45;   Gaps   3;

Qy   13 WPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTV 72

RESULT 6

```
Q8N7N8

ID   Q8N7N8;      PRELIMINARY;      PRT;   508 AA.
AC   Q8N7N8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ40785.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Trachea;
RA   Ohima A., Takahashi-Fujii A., Tanase T., Inose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa K., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kanda K.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK098104; BAC05233.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR000590; HMG_CoA synth.
DR   InterPro; IPR008260; P:acetyl-CoA synth.
DR   Pfam; PF01154; HMG_CoA_synt_1.
DR   PROSITE; PS01226; HMG_CoA_synt_AS.
SQ   SEQUENCE 508 AA; 56856 MW; 156DD32AF48084B4 CRC64;
```

Query Match 59.5%; Score 1494; DB 4; Length 508;
Best Local Similarity 59.4%;
Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps 1;

```
Qy    2 WPKDVGIIAMEVYVFPSQYVDQAELEEYDGVGAGKYTVGLSGQARMGFCSDREDINSLCLTV  61
Db   73 VQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGG   132
Qy   62 VQRLMERNGLSYESVGRLEVGTETIIDKSKSTKTVLMQLFEESGNTDVEGVDTTNACYGG   121
Qy  133 TAAVFNAVNWIESSSWD------------------------------------------G  150
Db  122 TAALFNAVNNWVESSSWDGRYALVVAGDIAVVATGSARPTGGAGAVAMLVGPNAPLAFERG  181
Qy  151 LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKIHAQWQKEGNDKD    210
Db  182 LRGTHMQHAYDFYKPDMVSEYPVVDGKLSIQCYLSALDQCYSVYVKNKIHARWQREGTEGR  241
Qy  211 FTILNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNS-IYSGLEAFGDVKLEDTYF  269
Db  242 CSLEDFGFMVFHSPYCKLVQKSLARLMLNDFLCHPSPNMESGPPFSGLEAFRDVKIEDTYF  301
Qy  270 DRDVEKAFMKASSELFSQKTKASLILVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIG  329
Db  302 DRDVEKAFMKASSELFDDKTKASLLISNQNGNMYTPSVYGCLASVLAQHTPQQLAGQRIG   361
Qy  330 VFSYGSGLAAATLYSLKTYQDDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRE 389
Db  362 VFSYGSGFAATLYSIKVTQDDATPGSALDKLVSSLCDLPARLDSRQKVSPGVFAETMKLRE  421
Qy  390 DTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTENDDTLDEGVGLVHSNIATE   449
Db  422 ETHHLANYIPQGSVDELFPGTWYLIRVDEKHRRQYARSMNDDRPLE--AGLVSSSMAAE   479
Qy  450 HIPSPAKKVPRLPATAAEPEAAVISNGEH  478
Db  480 HIPSPLKKMPRIPTTAGPEVVVMSNGDH   508
```

RESULT 7

```
Q9DBM4

ID   Q9DBM4;      PRELIMINARY;      PRT;   508 AA.
AC   Q9DBM4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   3-hydroxy-3-methylglutaryl-coenzyme A synthase 2.
GN   HMGCS2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Liver;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., King B.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409.685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
```

PRELIMINARY; PRT; 508 AA.

```
Qy    6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCFCTDREDI  65
Db   43 PLAKTDTWPKDVGILALEVVFPPAQVVDQTDLREKPNNVEAGKYTVGLGQTRMGFCSVQEDI 102
Qy   66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT   125
Db  103 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKVLMELFQDSGNTDIEGIDT   162
Qy  126 TNACYGGTAAVFNAVNWIESSSWD------------------------------------  149
Db  163 TNACYGGTGASLFNAANWMESSYWDGRYALVVCGDIAYPSGNARPTGGAGAVAMLIGPKA   222
Qy  150 ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKIHAQWQ    203
Db  223 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKKIQNQWK   282
Qy  204 KEGNDKDFTILNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK  263
Db  283 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARIMFNDFLSSSSDKQNNIYKGLEAFRGLK   342
Qy  264 LEDTYFDRDVEKAFMKASSELFSQKTKASLILVSNQNGNMYTSSVYGSLASVLAQYSPQQL  323
Db  343 LEETYTNKDVDKALLKASLDMFNQKTKASLIYLSTNNGNMYTSSLYGCLASLSHHSAQEL  402
Qy  324 AGKRIGVFSYGSGLAAATLYSLKTYQDDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAE 383
Db  403 AGSRIGAFSYGSGLAASFFSFRVSKDASPGSTLEKLVSSVSDLPKRLDSRRMSPEEFTE   462
Qy  384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRP  428
Db  463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLYLERVDEMHRKYARCP 507
```

RC TISSUE=Colon, and Liver;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK004865; BAB23626.1; -.
DR EMBL; BC014714; AAH14714.1; -.
DR EMBL; BC024744; AAH24744.1; -.
DR PIR; B57729; B57729.
DR MGD; MGI:101939; Hmgcs2.
DR GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR InterPro; IPR002810; HMG_CoA_synth.
DR InterPro; IPR000590; HMG_CoA_synth_AS.
DR Pfam; PF01154; HMG_CoA_synt; 1.
DR PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
SQ SEQUENCE   508 AA; 56822 MW; 5A20ACE918FF4758 CRC64;

Query Match   59.5%; Score 1493; DB 11; Length 508;
Best Local Similarity   59.4%; Pred. No. 2.3e-101;
Matches 276; Conservative 74; Mismatches 73; Indels 42; Gaps   1;

Qy   6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI 65
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43 PLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI 102

Qy   66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
          ||||||||||||||||||| :||||||||||||||||||||||||||||||||||||||
Db   103 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 162

Qy   126 TNACYGGTAAVFNAVNWIESSSWD------------------------------------ 149
          |||||||||||:||:|||||||||
Db   163 TNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNARPTGAGAVAMLIGPKA 222

Qy   150 ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ 203
                ||||||||||||:||||||||||:|| :||||||||||||||||||||: :|||
Db   223 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKKIQNQWK 282

Qy   204 KEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
          ||||||||| ||||||||||||||||||||||||||||||||||| |||||||||||||
Db   283 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLYKGLEAFRGLK 342

Qy   264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQL 323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   343 LEETYTNKDVDKALLKASLDMFNQKTKASLYLSTNNGNMYTSSLYGCLASLLSHHSAQEL 402

Qy   324 AGKRIGVFSYGSGGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAB 383
          |||||||||||||||||||||||||| || ||:||||:|||||:||:||||::|:|||
Db   403 AGSRIGAFSYGSGLAASFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462

Qy   384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARP 428
          ||:||||:|||:||:||| || ||||| || :| |||:| |
Db   463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRXYARCP 507

Query Match   59.4%; Score 1491; DB 11; Length 508;
Best Local Similarity   59.1%; Pred. No. 3.2e-101;
Matches 275; Conservative 75; Mismatches 73; Indels 42; Gaps   1;

Qy   6 PLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDI 65
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   43 PLAKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI 102

Qy   66 NSLCMTVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDT 125
          ||||||||||||||||||| :||||||||||||||||||||||||||||||||||||||
Db   103 NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGIDT 162

Qy   126 TNACYGGTAAVFNAVNWIESSSWD------------------------------------ 149
          |||||||||||:||:|||||||||
Db   163 TNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNARPTGAGAVAMLIGPKA 222

Qy   150 ------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQ 203
                ||||||||||||:||||||||||:|| :||||||||||||||||||||: :|||
Db   223 PLVLEQGLRGTHMENAYDFYKPNLASEYPLVDGKLSIQCYLRALDRCYAAYRKKIQNQWK 282

Qy   204 KEGNDKDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVK 263
          ||||||||| ||||||||||||||||||||||||||||||||||| |||||||||||||
Db   283 QAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARLMFNDFLSSSSDKQNNLYKGLEAFRGLK 342

Qy   264 LEDTYFDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQL 323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   343 LEETYTNKDVDKALLKASLDMFNQKTKASLYLSTNNGNMYTSSLYGCLASLLSHHSAQEL 402

Qy   324 AGKRIGVFSYGSGGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAB 383
          |||||||||||||||||||||||||| || ||:||||:|||||:||:||||::|:|||
Db   403 AGSRIGAFSYGSGLAASFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462

Qy   384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARP 428
          ||:||||:|||:||:||| || ||||| || :| |||:| |
Db   463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRXYARCP 507

[PRELIMINARY;   PRT;   465 AA.]

Qy   324 AGKRIGVFSYGSGGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAB 383
          |||||||||||||||||||||||||| || ||:||||:|||||:||:||||::|:|||
Db   403 AGSRIGAFSYGSGLAASFSFRVSKDASPGSPLEKLVSSVSDLPKRLDSRRRMSPEEFTE 462

Qy   384 NMKLREDTHHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARP 428
          ||:||||:|||:||:||| || ||||| || :| |||:| |
Db   463 IMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMHRRXYARCP 507

```
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila;
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.E.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003807; AAF50055.1; -.
DR   EMBL; AY051743; AAK93167.1; -.
DR   FlyBase; FBgn0010611; Hmgs.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   465 AA;  51176 MW;  57AA3A9FA4D424C8 CRC64;
```

```
Qy   Query Match             53.2%;  Score 1337;  DB 5;  Length 465;
     Best Local Similarity   57.3%;  Pred. No. 6e-90;
     Matches  263;  Conservative  56;  Mismatches  94;  Indels  46;  Gaps  3;

Qy    13  WPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTV  72
```

```
RESULT 10

ID   Q9NDA8          PRELIMINARY;       PRT;        457 AA.
AC   Q9NDA8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   3-hydroxy-3-methylglutaryl coenzyme A synthase.
GN   HMGS.
OS   Dendroctonus jeffreyi (Jeffrey pine beetle).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
OC   Phytophaga; Scolytidae; Dendroctonus.
OX   NCBI_TaxID=77162;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20500936; PubMed=11044666;
RA   Tittiger C., O'Keeffe C., Bengoa C.S., Barkawi L.S., Seybold S.J.,
RA   Blomquist G.J.;
RT   "Isolation and endocrine regulation of an HMG-CoA synthase cDNA from
RT   the male Jeffrey pine beetle, Dendroctonus jeffreyi (Coleoptera:
RT   Scolytidae).";
RL   Insect Biochem. Mol. Biol. 30:1203-1211(2000).
DR   EMBL; AF166002; AAF99580.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0003700; F:transcription factor activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000005; HTHAraC.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
SQ   SEQUENCE   457 AA;  50690 MW;  9AA9CD6B39E80F73 CRC64;
```

```
Qy   Query Match             46.4%;  Score 1164;  DB 5;  Length 457;
     Best Local Similarity   50.4%;  Pred. No. 3.2e-77;
     Matches  231;  Conservative  64;  Mismatches  109;  Indels  54;  Gaps  4;

Qy    11  ACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCM  70
```

```
Qy     15  KDVGIVALEIYFPSQYVDQAELEXYDGVDAGKYTIGLGQQAKMGFCTDREDINSLCMTVQ  74
           |||:||||:|||||||||:|||||||||||||||||||:||||:||||||||||||||:|
Db      6  KDVGILAMDIYFPPTCVLQDEELENHDGVSKGKYTIGLGQDSMAFCTEVEDVISMSLTVVK  65

Qy     75  NLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGTA  134
           |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6  KDVGILAMDIYFPPTCVLQDEELENHDGVSKGKYTIGLGQDSMAFCTEVEDVISMSLTVVK  65
```

RESULT 11

```
ID   Q8H051;
AC   Q8H051;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative hydroxymethylglutaryl coenzyme A synthase.
GN   OJ1263H11.4.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Currie J.,
RA   McCombie W.R., de la Bastide M., Spiegel L., Preston R.,
RA   Collura K., McCombie D.W.R., Nascimento L., Zutavern T., Balija V., Bell M.,
RA   Kirchoff K., Kuit K., Palmer L., Katzenberger F., Muller S., King L.,
RA   Baker J., Santos L., Miller B., Dedhia N.;
RA   Yang C., O'Shaugnessy A., Palmer L., Dedhia N.;
RT   "Rice Genomic Sequence.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC119980; AA015287.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
SQ   SEQUENCE   463 AA;  51502 MW;  60D021C5C48B99B6 CRC64;
```

```
Query Match             41.7%;  Score 1047;  DB 10;  Length 463;
Best Local Similarity   47.1%;
Matches  216;  Conservative   70;  Mismatches  111;  Indels   62;  Gaps    8;
```

```
Db      2  SAWPEDVGIIALEIYFPAQCYVDQTELEQVDGVSAGKYTVGLGQKKMGFCNDREDINSLCL  61

Qy     71  TVVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACY  130
           ||| :|   :   :  :|||||||||||||||||||||| ||||||| ||||||||||||
Db     62  TVVDNLLTRYQVKPQEIGRLEVGTETIIDKSKSVKTVLMQLFEPHGVTDIEGIDTTNACY  121

Qy    131  GGTAAVFNAVNWIESSSWDG-----------------------------------------  150
           ||                  :
Db    122  GGLPPSSTPSIAVESSSWNGRYALVVAADIAVAYAKGPARPTGGTCAVAMLVGPNAPLVVD  181

Qy    151  --LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGND  208
             |:   :|||||||||: ||||:|||||||||||: |||:|| |||| | :  ||:| ||
Db    182  RRLKSSFFGHAYDFYKPDLTSEYPVVDGKLSIQCYLEALDKCYQVYCQKTEKLLQ-----  236

Qy    209  KDFTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTY  268
           |   : :|  ::||:|||||||||| |:|||||    :   : :   ||| ||:|| |:|
Db    237  KPVSIESFPGILFHTPYCKLVQKSVGRLALNDFVRE---GKPELHPDLEKFANIQLKDSY  293

Qy    269  FDRDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSYVGSLASVLAQYSPQOLAGKRI  328
           |||||||||  ||  :| :||  |||::  :|||||||| ||||:| : : || :|||
Db    294  FDRDVEKAFLSASNAAFLQRTLPSLLIAANIGNMYTSSLYGGLVSYIVSRSLENLVDNRV  353

Qy    329  GVFSYGSGLAATLYSLKYTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLR  388
             ||||||||:: ||: :   |  | |||||||||||||||||||||||| |||:|  |
Db    354  AMFSYGSGLASSPYSITIKNNA----ALQTLKNNLSYEPLLAKRSKIAPEKFEATLELR  409

Qy    389  EDTHHLVNYIPQGSIDSLFEGTWVLVRVDEKHRRTYAR  426
           : :  :|| ||:|||| |||||:||:| |||||||  |
Db    410  QQSAHKVPYEPVGDISHFPGTFVLTKIDEQHRRVDR   447
```

RESULT 12

```
ID   Q9HE19;
AC   Q9HE19;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Putative 3-hydroxy-3-methylglutaryl coenzyme A synthase.
GN   HMGS.
OS   Phycomyces blakesleeanus.
OC   Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
OC   Phycomyces.
OX   NCBI_TaxID=4837;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL1555;
RA   Ruiz-Albert J., Cerda-Olmedo E., Corrochano L.M.;
RT   "Genes for the metabolism of 3-hydroxy-3-methylglutaryl coenzyme A in
RT   the fungus Phycomyces.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ297414; CAC18553.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
SQ   SEQUENCE   450 AA;  50136 MW;  92B6E495039F0967 CRC64;
```

```
Query Match             41.6%;  Score 1044;  DB 3;  Length 450;
Best Local Similarity   47.5%;
Matches  217;  Conservative   68;  Mismatches  116;  Indels   56;  Gaps    6;
```

```
Db     66  SLLENYKIDPKCIGRLEVGSETVIDKSKSIRTKWLMQIFBECGNTDIEGVDSSNACYGGTA  125

Qy    135  AVFNAVNWIESSSWDG----------------------------LR  152
           |:   ||:||||:::|                            :
Db    126  ALFNCVNMVFESSNWDGRYGLVVCTDSAVAEGPARPTGGAAAIAMLIGPNAPVSFESKYR  185

Qy    153  GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKDFT  212
           |:||::|||||||| |||| :|||||| |||| ||| ||.|:||| :   ||| |
Db    186  GSHMAHVYDFYKPDLASEYPVVDGKLSQTCYLMALDSCYNVFCKKYE---KQFS  240

Qy     13  WPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQQAKMGFCTDREDINSLCMTV  72
            | |||| ||| :||| |  :: |:|||: |||||||||||  : ||  |:|||| |||| |
Db      7  YPENVGILALEMYFPSRCVEQOTAMEVYDGVSTGKYTIGLGQDKWAFIDDREDIQSICLTA  66

Qy     73  VQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGG  132
           | |||| |:   :|||||||||||||||||||||| ||:||:  |  ||||||||||||
Db     67  VHNLMEKYNIAYTDIGRLEVGTETIIDKSKSVKTILMTLFAEHGNYEIEGIDTTNACYGG  126

Qy    133  TAAVFNAVNWIESSSWD------------------------------------------  150
           ||
Db    127  FSAFSNAVNWIESSSWDGRVAIVIAGDLALYASGAARPTSGAGVAMLVGKDAPIVLGRG  186

Qy    151  LRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWQKEGNDKD  210
           ||.| |:|||||||||||:|||||||| | |:||: |:|||..| |: :: :    |||
Db    187  LRYTYMDHAYDFYKPDMHSEYPIVDGKFPSNVCYIRAFDACYNNYMKRLSKTIKKDVT--  243

Qy     2   SAMPEDVGIIALEIYPPACYVDQTELEQVDGVSAGKYTVGLGQKKMGFCNDREDINSLCLCL  61
```

```
Qy  211 FTLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFD 270
            ::: ||||:||||:|::  ::| ::||  |||::::| :||:|  :::::|:||: :|:
Db  244 -SMEDVDYVCHSPYAKLVTKSFARSAYNDFMADK---KNPKYAALAPFEELAYBASLEN 299

Qy  271 RDVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQOLAGKRIGV 330
            ||:||| | : |::: | |||  ||   |||| :.||| |||:|::|:|   ||||:
Db  300 RDLEKATATLTKAGVAQKVGPAAYAPKQIGNMYAGAVWAGLASLVSEVEVDTLIKDKRVLL 359

Qy  331 FSYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLRED 390
            :||||||| :: |::|     |: || | : |  |:|||| || .|    :|  ::||:
Db  360 YSYGSGLAASMVSFRVV-----GSTAD-IKAKL-NLRQRLDARTHSKPRAFARAMQIREN 412

Qy  391 THHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARR 427
            || :.| ||:|:|.: . | :| |::|:| || |:|
Db  413 THNACNYNPIGSLEHIAPGAYYVDKIDDKWRRFYKRK 449
```

RESULT 13
Q8C5F4

```
ID   Q8C5F4;      PRELIMINARY;      PRT;      268 AA.
AC   Q8C5F4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hydroxymethylglutaryl-CoA synthase.
GN   HMGCS1 OR B13003ZCO6RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RA   The FANTOM Consortium.
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK078743; BAC37373.1; -.
DR   MGD; MGI:1075992; Hmgcs1.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
SQ   SEQUENCE   268 AA;  9F6A7B8AFD4317F4 CRC64;

Query Match              41.0%;  Score 1029.5;  DB 11;  Length 268;
Best Local Similarity    72.7%;  Pred. No 1.1e-67;
Matches  200;  Conservative  10;  Mismatches  14;  Indels  51;  Gaps  2;

Qy    1 MPGSLPLNAERACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCT 60
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1 MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQARMGFCT 60

Qy   61 DREDINSLCMTVQNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDI 120
          |||||||||:||||:||| :|:||||||||||||||||||||||| :||||||||||||
Db   61 DREDINSLCLTVVQKLMERHSLYDCIGRLEVGTETIIDKSKSVKSNLMQLFEESGNTDI 120

Qy  121 EGIDTTNACYGGTAAVFNAVNWIESSSWD------------ 149
          ||||||||||||||||||||||::|||||
Db  121 EGIDTTNACYGGTAAVFNAVNWVESSSSWDGRYALVVAGDIAIYATGNARPTGGVGAVALL 180

Qy  150 ------------GLRGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKI 198
                      ||||||||||||:|||||||||||:||||||||||||||||||| ||:|
Db  181 IGPNAPLIFDRGLRGTHMQHAYDEYKPDMLSEYPVVDGKLSIQCYLSALDRCYSVRKKI 240

Qy  199 HAQWOKEGKGNDKDFTLNDFGFMIFHSPYCKLVQKSL 233
          :||| |:
Db  241 RAQWQKGG-------FSFSQFSCTRCWCLEGSV 266
```

RESULT 14
Q944F8

```
ID   Q944F8;      PRELIMINARY;      PRT;      464 AA.
AC   Q944F8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hydroxymethylglutaryl coenzyme A synthase.
OS   Hevea brasiliensis (Para rubber tree).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids I; Malpighiales; Euphorbiaceae; Crotonoideae; Micrandreae;
OC   Hevea.
OX   NCBI_TaxID=3981;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. RRIM600;
RA   Hallahan D.L., Keiper-Hrynko N.M.;
RT   "Genes involved in the biosynthesis of isopentenyl diphosphate in the
RT   rubber tree Hevea brasiliensis.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF429389; AAL18930.1; -.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   InterPro; IPR000590; HMG_CoA_synt_AS.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS01226; HMG_COA_SYNTHASE; 1.
SQ   SEQUENCE   464 AA;  980060294CB85187 CRC64;

Query Match              40.3%;  Score 1013;  DB 10;  Length 464;
Best Local Similarity    44.2%;  Pred. No. 4.2e-66;
Matches  207;  Conservative  86;  Mismatches  121;  Indels  54;  Gaps  8;

Qy   15 KDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVVQ 74
          |:||| | : | |:  | :|| |  ||.  |||||||||| |  |||:|||||| |:.
Db    3 KNVGILAVDIYFPPTFVQQEALEAHDGASKGKYTIGLGQDCMAFCTEVEDVISMSLTAVT 62

Qy   75 NLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGTA 134
          :||| :|:|||||||| |:||:||||||:|| :| :|||| ||||||:|:|||||||||
Db   63 SLLDKYNIDEKQIGRLEVGSETVIDKSKSIKTFLMQIFEPKFRGNTDIEGVDSTNACYGGTA 122

Qy  135 AVFNAVNWIESSSWD------------------------------ 152
          |:||.||| ||||||
Db  123 ALFNCVNWVESSSWDGRYGLVVCTDSAVYAEGPARPTGGAAAIAILVGPDAPIAFESKFR 182

Qy  153 GTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWOK-EGNDKDF 211
          |:||.||||||||::|:|||:|||||| : | ::|| ||.| |     |::| : .|:.|
Db  183 GSHMSHAYDFYKPNLASEYPVVDGKLSQTCVIMALDSCYKHFC----AKYEKFEG--KQF 236

Qy  212 TLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDR 271
          ::.  :|:.|||||.|||||| ||:: |||:.. .  :::| ||||:  :::  ::.:.|
Db  237 SISDAEYFVFHSPYNKLVQKSFARLVFNDFVRNASSIDETAKEKLAPFSNLSGDESYQNR 296

Qy  272 DVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRIGVF 331
          |:||. : |:   | .  | :::: .:.:|||| |:| ::| |:|   :|::|||| :.|
Db  297 DLEKVSQQVAKPLYDAKKPTTLIPKQVGNMYTASLYAAFASLL-HSKHTELAGKRVTLF 355

Qy  332 SYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDT 391
          ||||||:||::|| : :   | ::|:   |:: ::   | :     :  ::::|||:| .
Db  356 SYGSGLTATMFSLRLHEGQHPFS-LSNI-ASVMNVAGKLKARHELPPEKFVDIMKLMEHR 413

Qy  392 HHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTPNDDTLDEGV 439
            | :. |: :  : : | .|| .||: | || |:|       ::.|
Db  414 YGAKDFVTSKDCSLLASGTYYLTEVDSLYRRFYAQKAVGN--TVENGL 459
```

RESULT 15
P93773

```
ID   P93773;      PRELIMINARY;      PRT;      474 AA.
AC   P93773;
```

```
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   3-hydroxy-3-methylglutaryl-CoA-synthase.
OS   Pinus sylvestris (Scots pine)
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX   NCBI_TaxID=3349;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Needle;
RX   MEDLINE=97214637; PubMed=9061017;
RA   Wegener A., Gimbel W., Werner T., Hani J., Ernst D., Sandermann H.;
RT   "Molecular cloning of ozone-inducible protein from Pinus sylvertris L.
RT   with high sequence similarity to vertebrate 3-hydroxy-3-methylglutaryl-
RT   CoA-synthase.";
RL   Biochim. Biophys. Acta 1350:247-252(1997).
DR   EMBL; X96386; CAA65250.1; -.
DR   PIR; T09688; T09688.
DR   GO; GO:0004421; F:hydroxymethylglutaryl-CoA synthase activity; IEA.
DR   GO; GO:0006084; P:acetyl-CoA metabolism; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   InterPro; IPR008260; HMG_CoA_synth.
DR   Pfam; PF01154; HMG_CoA_synt; 1.
DR   PROSITE; PS00014; ER_TARGET; 1.
SQ   SEQUENCE   474 AA;  52998 MW;  3ECB48E5F23D37BA CRC64;

Query Match             40.2%;  Score 1009.5;  DB 10;  Length 474;
Best Local Similarity   43.5%;  Pred. No. 7.8e-66;
Matches 208;  Conservative  76;  Mismatches 141;  Indels  53;  Gaps   7;

Qy      14 PKDVGIVALEIYFPSQYVDQAELEKYDGVDAGKYTIGLGQAKMGFCTDREDINSLCMTVV 73
           |:||| |||||||:||:|||:||||||||||||||||| |:||| || :||||:|  ||
Db       5 PENVGILAMEIYFPTTCVQQEDLETFDGVSKGKYTIGLGQDCMTFCCTDLEDVISMSLTAV 64

Qy      74 QNLMERNNLSYDCIGRLEVGTETIIDKSKSVKTNLMQLFEESGNTDIEGIDTTNACYGGT 133
           :|:|:   |||||||:||||||||||||:|||||:|||||||||||||||||||||||||
Db      65 TSLLEKYEIDPKOIGRLEVGSETVIDKSKSIKTWLMHIFEKCGNTEIEGGVDSTNACYGGT 124

Qy     134 AAVFNAVNWIESSSWDG------------------------------------------L 151
           |||||||||||||||||
Db     125 AALFNCINWIESSSWDGRYGLVVATDSAVYAEGAARPTCGAAAVAMLIGPNAPIATESKY 184

Qy     152 RGTHMQHAYDFYKPDMLSEYPIVDGKLSIQCYLSALDRCYSVYCKKIHAQWOKEGNGNDKDF 211
           |||||:|||||||||::||||||||||||:|||||||:||||:||||:|   :||||:|
Db     185 RGTHMAHVYDFYKPNLASEYPVVDGKLSQTCYLMALDSCYKRPCNKFE----KEEG--RQF 239

Qy     212 TLNDFGFMIFHSPYCKLVQKSLARMLLNDFLNDQNRDKNSIYSGGLEAFGDVKLEDTYFDR 271
           ||:|| ||| :||:|| |::  |: |  |:|:|||||::||:| :|| |||||||:|:
Db     240 SLLDTDYIAFHSPYNKLVQKSFGRLLFNDFSRHARSVGKDAQEKLEPPAGLSEQDSYNSR 299

Qy     272 DVEKAFMKASSELFSQKTKASLLVSNQNGNMYTSSVYGSSLASVLAQYSPQQLAGKRIGVF 331
           |:|||||:|||||| :|||||:|||  ||:||||||||||||||||:||||||||||:|:
Db     300 DLEKVSGQLAKPLYDAKIQPSTLLPKQVGNMYTASLYAALASII-HNKHTTLDGQRVNMF 358

Qy     332 SYGSGLAATLYSLKVTQDATPGSALDKITASLCDLKSRLDSRTGVAPDVFAENMKLREDT 391
           ||||||||:||:||:|| |:|:|||:||:||:|||||||||||:||| ||:| ||||:||
Db     359 SYGSGLIASTLFSPKIREGQPP-FTLSNIT-EVMDVQNKLDSRHEFLPEDFVENLKRMETL 416

Qy     392 HHLVNYIPQGSIDSLFEGTWYLVRVDEKHRRTYARRPTENDDTLDEGV---GLVHSNI 446
           :  :   |:::| || ::| |:||:||  :|::|||| ||| :     :   :|: :||
Db     417 YGAKDFVSTSQLSLLRPGAFYLTKVDSMYRRFYSRKVISAGDNPEKSKLANGTTHDEL 474
```

Search completed: June 24, 2004, 13:19:30
Job time : 88 secs

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 24, 2004, 13:28:14 ; Search time 2861 Seconds
                                            (without alignments)
                                            4989.207 Million cell updates/sec

                        GenCore version 5.1.6
                        Copyright (c) 1993 - 2004 Compugen Ltd.

Title: US-10-622-516-2
Perfect score: 2511
Sequence: 1 MPGSLPLNAEACWPKDVGIV.............PRLPATAAEPEAAVISNGEH 478

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext 0.5
                 Ygapop 10.0 , Ygapext 0.5
                 Fgapop 6.0 , Fgapext 7.0
                 Delop   6.0 , Delext 7.0

Searched: 2751289 seqs, 1491090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1p
-Q=/cgn2_1/USPTO_spool_p/US10622516/runat_23062004_162633_660/app_query.fasta_1.647
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10622516 @CGN 1_1 5180 @runat_23062004_162633_660 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   EST:*
                 1: em_estba:*
                 2: em_esthum:*
                 3: em_estin:*
                 4: em_estmu:*
                 5: em_estov:*
                 6: em_estpl:*
                 7: em_estro:*
                 8: em_htc:*
                 9: gb_est1:*
                 10: gb_est2:*
                 11: gb_htc:*
                 12: gb_est3:*
                 13: gb_est4:*
                 14: gb_ests:*
                 15: em_estun:*
                 16: em_estom:*
                 17: em_est5:*
                 18: em_gss_hum:*
                 19: em_gss_inv:*
                 20: em_gss_pln:*
                 21: em_gss_vrt:*
                 22: em_gss_fun:*
                 23: em_gss_mam:*
                 24: em_gss_pro:*
                 25: em_gss_rod:*
                 26: em_gss_phg:*
                 27: em_gss_vrl:*
                 28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
1	2348	93.5	3222	11	AK045094	AK045094 Mus muscu
2	2348	93.5	3286	11	AK044835	AK044835 Mus muscu
3	2348	93.5	3466	11	AK031297	AK031297 Mus muscu
4	1513	60.3	1527	29	AY416670	AY416670 Homo sapi
5	1493	59.5	1527	29	AY416672	AY416672 Mus muscu
6	1493	59.5	3287	11	AK004865	AK004865 Mus muscu
7	1491	59.4	2524	11	AK004902	AK004902 Mus muscu
8	1382	55.0	1083	13	BX441322	BX441322 Mus muscu
9	1285	51.2	1420	29	AY416671	AY416671 Pan trogl
10	1263	50.3	998	12	BM474567	BM474567 AGENCOURT
11	1263	50.3	1201	13	BX419944	BX419944 AGENCOURT
12	1255.5	50.0	905	13	BU508388	BU508388 AGENCOURT
13	1247	49.7	789	14	CA512392	CA512392 UI-M-GMO-
14	1216	48.4	967	13	CD354310	CD354310 UI-M-GMO-
15	1210	48.2	780	14	BU149049	BU149049 AGENCOURT
16	1206	48.0	960	12	CA316122	CA316122 UI-M-FW0-
17	1197	47.7	914	13	BQ009460	BQ009460 AGENCOURT
18	1196	47.6	914	13	BU149022	BU149022 AGENCOURT
19	1173.5	46.7	964	13	BU514917	BU514917 AGENCOURT
20	1171.5	46.7	877	9	AU132233	AU132233 AGENCOURT
21	1149	45.8	810	13	CD354238	CD354238 UI-M-GMO-
22	1147	45.7	937	13	BX441044	BX441044 AGENCOURT
23	1146.5	45.7	1106	14	CD505314	CD505314 UI-M-GMO-
24	1143	45.5	1156	12	BM461079	BM461079 CDA73-A07
25	1138	45.3	1118	14	CD503546	CD503546 AGENCOURT
26	1136	45.2	914	13	BU941319	BU941319 AGENCOURT
27	1135	45.2	718	12	BM963690	BM963690 UI-M-EQ0-
28	1134	45.2	1053	13	BU162936	BU162936 AGENCOURT
29	1133.5	45.1	1068	14	BU503357	BU503357 CDA61-H01
30	1133.5	45.1	881	13	BX853811	BX853811 AGENCOURT
31	1132	45.1	881	13	CF169002	CF169002 UI-M-GM0-
32	1117	44.5	866	13	BX742538	BX742538 AGENCOURT
33	1115	44.4	725	14	CD354751	CD354751 UI-M-GMO-
34	1112	44.3	674	14	CA873594	CA873594 K0926C09-
35	1111	44.2	674	14	CF170711	CF170711 B0832B04-
36	1111	44.2	950	14	CF412994	CF412994 CH3#083_D
37	1108	44.1	682	14	CB288167	CB288167 CMD75_F05
38	1104	44.0	669	14	CA874751	CA874751 K093AF10-
39	1101	43.8	885	12	BG681747	BG681747 602627868
40	1099	43.8	823	12	BI871842	BI871842 60399698I
41	1099	43.8	1201	13	BX445244	BX445244 AGENCOURT
42	1093.5	43.5	756	14	CF737178	CF737178 UI-M-HD0-
43	1091	43.4	703	9	AU127534	AU127534 UI-M-GM0-
44	1089	43.4	769	14	CD353582	CD353582 UI-M-GMO-
45	1087	43.3	657	14	CA874680	CA874680 K093AB05-

ALIGNMENTS

RESULT 1
AK045094
LOCUS   AK045094   3222 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION   Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
                enriched library, clone:B1003z2C06 product:pre B-cell leukemia
                transcription factor 1, full insert sequence.
ACCESSION   AK045094
VERSION   AK045094.1 GI:26337068
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530013
  PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6   (bases 1 to 3222)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,Y., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES   Location/Qualifiers
     source   1..3222
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130032C06"
                     /db_xref="MGI:2410453"
                     /clone="B130032C06"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /tissue_type="parthenogenote"
                     /dev_stage="9.5 days embryo"
                     /db_xref="taxon:10090"
     CDS   62..1624
                     /note="unnamed protein product; pre B-cell leukemia
                     transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
                     match=124)
                     /codon_start=1
                     /evidence=BLASTN, 99%, match=124)
                     /protein_id="BAC32218.1"
                     /db_xref="GI:26337069"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFSQYDQAELEKYDGVDA
                     GKYIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNVANVNWVBSSSWDGRYALVVAGDI
                     AIVATGNARPTGGVGAVALLIGPNAPLIFDRGLRGTEMQHAYDFYXPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQRDKRNSIYSGLEAPGDVKLEDTYDTEDKAFARMKASSELFNQKTKAS
                     LLVSNQNGNMYTSSVVGSLASVLAQYSPQOLAGKRVGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRICVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                     ESESAVISNGHH"
     polyA_signal   3208..3213
                     /note="putative"
     polyA_site   3222
                     /note="putative"
ORIGIN
Alignment Scores:
Pred. No.:
Score: 3.34e-258   Length: 3222
Percent Similarity: 89.42%   Matches: 452
Best Local Similarity: 86.92%   Conservative: 13
Query Match: 93.51%   Mismatches: 13
                                                    Indels: 42
DB: 11   Gaps: 1

US-10-622-516-2 (1-478) x AK045094 (1-3222)

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCCTTGCTCGGCCAAAAGATGTGGGAATCGTT   121

Qy   21 AlaLeuGluIleTyrPheSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
          |||||||||||||||||||||||||:::|||||||||||||||||||||||||||:::
Db   242 GCCCTTGAAATCTACTTTCCTTCTCAAATARGTCGATCAAGCTGAGTTGGAAAAATACGAT   181

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaArgMetGlyPheCysThr   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 GGTGTAGATGCYGGAAAGTATACCATCGGCCTGGGCCAGGCGCAGGATGGGCCTTCTGCACG   241

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
          |||||||||||||||:|:|||||||||||||||||||||||||||||||||||||:::
Db   302 AGCCTTCCTATGATTGCATTGGGCGCGGCTAGAAGTTGGACAGAGACATCATCGACAAA   361

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluGlyThrGluThrIleIleAspLys   100
          :::|||||||||||||||||||||||||||||||||||||||||||||||||:||||
Db   302 AGCCTTCCTATGATTGCATTGGGCGCGGCTAGAAGTTGGACAGAGACATCATCGACAAA   361

Qy   101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle   120
          ||||||||||||:||||||||||||||||||:||||||||||||||||||||||||||||
Db   362 TCGAAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCTGGGGAATACAGATATA   421

Qy   121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaVal   140
          ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 GAAGGAATAGATACAACCAATGCATGCTATGGGGGCACAGCTGCAGTTCTTCAATGCCGTG   481

Qy   141 AsnTrpIleGluSerSerSerTrpAsp----------------------------------   149

```
Qy    482 AACTGGGGTCGGATCCAGCTCTTCTTGGGGATGGACGATATGCTCTGGTAGTTGCAGGAGACATT 541
          ||::||||||:|||||||||||||||||||||||||
Db    149 ---------------------------------------------------------------- 149

Qy    542 GCTATATATGCCACAGGAAATGCCAGACCTACAGGTGGAGTTGGAGCTGTGGGCCCTGCTA 601
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 ---------------------------------------------------------------- 149

Qy    602 ATTGGGGGCCAAAGGCTCTCTAATTTTTGACCGAGGGCTTCGGTGGGACACATGCAGCAT 661
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 --------------------GlyLeuArgGlyThrHisMetGlnHis 158

Qy    662 GCCTATGACTTTTACAAGCCTGACATGCTCTCCGAGTACCCTATCGTGGCGGGAAAGCTC 721
          ||||||||||||||||||||||||||||||||||||||||::|||||||||||||||
Db    159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAlaGlyLysLeu 178

Qy    722 TCCATACAGTGCTACTACCTCAGCCGCGCCCTGACCGCGTACTCGGTCTACTGCGCAAAGATC 781
          ||||||||||||||||||||||||||||||||||||||
Db    179 SerIleGlnCysTyrTyrLeuSerArgAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198

Qy    782 CCTGCCCAGTGGCAGAAAGGAGAAGGAAATGATCCATTTACCCTGAATGATTTTGGCTTC 841
          ||||||||||||||||||||||||||||||||||||||
Db    199 HisAlaGlnInTrpGlnLysGlyGluGlyAsnAspProPheThrLeuAsnAspPheGlyPhe 218

Qy    842 ATGATCTTTCACTCACCATATTGTAAACTGGTCTGCAGAAATCTCTAGCTCGGAGTTCCTG 901
          ||||||||||||||||||||||||||||||||||||||||
Db    219 MetIlePheHisSerProThrCysLysLysGleuValGlnIleValSerGlnLeuLeu 238

Qy    902 AATGACTTTCTTAATGATCAAAAACAGAGATAAAAAACAGTAATTTACAGTGGACTGGAAGGC 961
          ||||||||||||||||||||||||||||||||||||||
Db    239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerileTyrSerGlyLeuGluAla 258

Qy    962 TTTGGGCAGTTAAATTAGAAGATTACTTTGACACAGATGTAGAAAAGGCATTTATG 1021
          ||||||||||||||||||||||||||||||||||||||
Db    259 PheGlyAspValLysLeuGluAspThrTyrPheAspAspValGlyLysAlaPheMet 278

Qy   1022 AAGGCTAGTTCTGGAGCTATTCAACCAGAAAACAAAGGGCTCTTTGCTTGTGTCTAATTCAG 1081
          ||||||||||||||||||||||||||||||||||::|||
Db    279 LysAlaSerSerGlyLeuPheSerGlnLysGlyThrLysAlaSerLeuLeuValSerAsnGln 298

Qy   1082 AATGGAAATATGTACACAATCCCTGTCTAGGGTTCCCTCTGGCCTCGGCACAGTAC 1141
          |||||||||||||||||||||||||||||||||||||
Db    299 AsnGlyLysAsnMetLeuTyrThrSerSerArgValThrGlySerLeuAlaSerValLeuLeuGlnTyr 318

Qy   1142 TCACCTCAGCACAGTTGGGCGGAAGGGTTGCGAGTGTTCTTCTACGGTTCTTCTGGCTTGGCT 1201
          |||||||||||||||||||||:||::|||||||||||||||||
Db    319 SerProGlnGlnLeuAlaGlyIleGlyLysGlyValPheSerGlyTyrGlySerGlyLeuAla 338

Qy   1202 GCCACGCGTACTCCCTTAAAGTACAACAAGAGCACCGATCCTGCCCCTTGATAAA 1261
          ||||||||||||||||||||||||||||||||||::|
Db    339 AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspGlyLys 358

Qy   1262 ATAACAGCAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGGCACCG 1321
          ||||||||||||||||||||||||||||||||||
Db    359 IleLeuThrAlaSerLeuCysAspLeuLysSerArgGleuAspSerArgThrGlyValAlaPro 378

Qy   1322 GAGTCTTGCTGAAAACATGAAGCTCAGAGGAGCACACATCTACGGTTCTTCACGGCTTGGCT 1381
          ||||||||||||||||||||||:||||||||::||||||||
Db    379 AspValPheAlaGluLeuAsnMetLysLeuArgGlnAspThrHisHisileLeuValAsnTyrIle 398

Qy   1382 CCCAAGCGTGTTCAATAGACTCACTCTTTGAAGGAACGTGGTATCTGGTCCGAGTGGATGA 1441
          ||||:||||||||||||||||||||||||||||||||
Db    399 ProGlnGlySerLeuAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418

Qy   1442 AAACACAGAAGGACTTACCCCGGCGCCCTTCACAAATGACACCAGTTTGGATGAAGGA 1501
          ||||||||||||||||||||||||:||||||||||||||
Db    419 LysHisArgArgThrTyrAlaArgProThrProAsnAspAspThrLeuAspGluGly 438

Qy   1502 ATGGGGCTCGTGCATAGTAACACAGCAACAGAGCCATAATTCCAAGCCCTGCTAAGAAAGTG 1561
          |||||||||||::||||||||||||||||||||||||||
Db    439 ValGlyLeuValHisSerAsnIleAlaThrGluHisileProSerProAlaLysLysVal 458

Qy   1562 CCAAGACTCCCTGCAACCTCCGGCGCCGAAATCTGAATCAGCGTGTCATCAGTAACGGGGGGAGCAC 1621
          ||:||||||||||||||||||||||||||::|||||||||||||||
Db    459 ProArgLeuProAlaThrAlaAlaGluProGluAlaAlaValIleSerAsnGlyGluHis 478
```

Db   1562 CCAAGACTCCCTGCAACCTCCGGCGCCGAAATCTGAATCAGCGTGTCATCAGTAACGGGGGGAGCAC 1621

RESULT 2
AK044835

LOCUS   AK044835   3286 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130007E24 product:pre B-cell leukemia
            transcription factor 1, full insert sequence.
ACCESSION   AK044835
VERSION   AK044835.1 GI:26336856
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3286)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,

```
COMMENT      Fax:81-45-503-9216)
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in Riken.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Tissues were provided by Dr. Tomohiro Kono (Department of Animal
             Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
             Kanagawa Prefecture, Japan) whose assistance we gratefully
             acknowledge.
             Please visit our web site for further details.
             URI:http://genome.gsc.riken.go.jp/
             URI:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3286
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130007E24"
                     /db_xref="MGI:2410190"
                     /db_xref="taxon:10090"
                     /clone="B130007E24"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /tissue_type="parthenogenote"
                     /dev_stage="9.5 days embryo"
     CDS             123..1685
                     /note="unnamed protein product; pre B-cell leukemia
                     transcription factor 1 (MGD|MGI:97495, GB|NM_008783,
                     evidence: BLASTN, 99%, match=124)
                     /codon_start=1
                     /protein_id="BAC32112.1"
                     /db_xref="GI:26336857"
                     /translation="MPGSLPLNAEACWPKDVGIVALEIYFPSQYVDQAELEKYDGVDA
                     GKYTIGLGQARMGFCTDREDINSLCLTVVQKLMERHSLSYDIGRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTIDIEGIDTTNACYGTAAVFNAVNMWVESSNDGRYALVVAGDI
                     AIYATGNARPTGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                     KLSIQCYLSALDCYSVYRKIRAQWQKEGKDKDPTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQNRDKNSIYSGLEAFGDVKLEDTYFDRDVEKAFMKASSELFNQKTKAS
                     LIVSNQNGNMYTSSVYGSLASVLAQYSPQQLAGKRVGVFSYGSGLAATLYSLKVTQDA
                     TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWLIVRDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSPAKKVPRLPATSA
                     ESESAVISNGEH"
     polyA_signal    3272..3277
                     /note="putative"
     polyA_site      3286
                     /note="putative"
ORIGIN
```

Alignment Scores:
Pred. No.:

Score:	3.45e-258	Length:	3286
	2348.00	Matches:	452
Percent Identity:	89.42%	Conservative:	13
Percent Similarity:	86.92%	Mismatches:	13
Best Local Similarity:	93.51%	Indels:	42
Query Match:	11	Gaps:	1
DB:			

```
US-10-622-516-2 (1-478) x AK044835 (1-3286)

Qy    1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123  ATGCCTGGGTCACTTCCTTTGAATGCAGAGGCCTGCTGGCCAAAAGATGTGGAATCGTT  182

Qy   21  AlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183  GCCCTTGAAATCTACTTTCCTTCCCAATATGTCGATCAAGCTGAGTTGGAAAAATACGAT  242

Qy   41  GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaArgMetGlyPheCysThr   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243  GGTGTAGATGCTGGAAAGTATACCATCGGCCTGGGCCAGGCCAGGATGGGCTTCTGCACG  302

Qy   61  AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
         |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303  GATCGTGAAGACATCAACTCTTTGCCTGACTGTGGTTCAGAAACTGATGGAGAGACAT   362

Qy   81  AsnLeuSerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLys  100
         :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363  AGCCTTTCCTATGATTGCATTGGGCGGCCTAGAAGTTGGAACAGAGACAATCATCGACAAA  422

Qy  101  SerLysSerValLysSerThrAsnLeuMetGlnLeuPheGluGluSerGlyAsnThrAspIle  120
         |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
Db  423  TCGAAATCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTCTGGGAATACAGATATA  482

Qy  121  GluGlyIleAspThrThrAsnAlaCysTyrGlyThrAlaAlaValPheAsnAlaVal  140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483  GAAGGGATAGATACAACAAATGCATGCTATGGGACAGCACCTGCAGTTCTTCAATGCCGTG  542

Qy  141  AsnTrpIleGluSerSerSerTrpAsp-                                149
         |||||||||||||||||||||||||||
Db  543  AACTGGGTCGAATCCAGCTCTTGGGATGGAC--------------------------  602

Qy  149  ------                                                      149
Db  603  GCTATATATGCCACAGGAAATGCCAACCTACAGGGTGGAGTTGGAGTGTGCCCTGCTA  662

Qy  150  ------GlyLeuArgGlyThrHisMetGlnHis                            158
                    |||||||||||||||||||||||||
Db  663  AATTGGGCCAAAACGCTCCTCTAATTTTGACCGAGGGCTCCGGTGGGACACACATGCAGCAT  722

Qy  159  AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu  178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  723  GCCTATGACTTTTACAAGCCTGACATGCTCTCCGAGTACCCTGTCGACGGAAAGCTC  782

Qy  179  SerIleGlnCysTyrLeuSerAlaLeuAspAspCysTyrSerValTyrCysLysLysIle  198
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  783  TCCATACAGTGCTACCTCAGCGCCCTGGACGACTGCTATTCTGTCTACTGCAAAAAGATC  842

Qy  199  HisGlnGlnThrTrpGlnLysGluGlyLysAspAspPheAsnPheGlyPhe         218
         |||||||||||||||||||||||||||||||||||||||||:||||||||||
Db  843  CGTGCCCAGTGGCAGAAAGAGGGGAAAGATAAAGATCCTAACCTGAATGATTTTGGCTTC  902

Qy  219  MetIlePheHisSerProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuLeu  238
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  903  ATGATCTTTCACTCACCATATTGTAAACTGGTGCAGAAAATCTCAGCTCGGATGTTCCTG  962

Qy  239  AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla  258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  963  AATGACTTTCTTAATGATCAAAACAGAGATAAAAACAGTATTTACAGTGGACTGGAAGCC  1022

Qy  259  PheGlyAspValLysLeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMet  278
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1023  TTTGGGGACGTTAAATTAGAAGACACTTACTTTGACAGAGATGTAGAAAAGGCATTTATG  1082

Qy  279  LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuValSerAsnGln    298
         |||||||||||||||||||||||||||||||:||||||||||||||||||||
Db 1083  AAGGCTAGTTCTGAGCTATTCAGCCAGAAAACGAAAGCGTCTTTGGTCAGCAATCAG  1142

Qy  299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr  318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1143  AATGGAAATATGTACACAAGCTCTGTCTATGGTTCCCTGGCTTCTGTCCTGGCACAGTAC  1202

Qy  319  SerProGlnGlnLeuAlaGlyLysArgValGlyValPheSerTyrGlySerGlyLeuAla  338
         |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db 1203  TCACCTCAGCAGTTGGCAGGGAAGAGGGTTGGAGTGTTCTCTTACGGTTCTGGCTTGGCT  1262

Qy  339  AlaThrLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLys  358
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1263  GCCACGCTGTACTCCCTTAAAGTCACACAAGACGCCACACCAGGATCCGCCCTTGATAAA  1322

Qy  359  IleThrAlaSerLeuCysAspLeuLysSerArgLeuAspSerArgThrCysValAlaPro  378
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1323  ATAACAGCAAGTTTATTGGACCTTAAATCAAGGCTTGACTCGAGACTTGGCACCG  1382

Qy  379  AspValPheAlaGluAsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIle  398
         |||||||||||||||||||||||||||||||||||||||||||||||||:::
Db 1383  GATGTCTTTGCTGAAAACATGAAGCTCAGAGGAGGAGGACACATCACTTAGCCAACTATATT  1442
```

```
Qy   399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1443 CCCCAGTGTCAATAGACTCACTCTTTGAAGGAACGTGGTATCTGGTCGACGAGTGGATGAA 1502

Qy   419 LysHisArgGlyArgThrTyrAlaArgArgProThrProAsnAspAspThrLeuAspGlyLy 438
         ::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1503 AAACACAGAAGGACTTACGCCCGGCCCCTTCACAAATGACCACAGTTTGGATGAAGGA 1562

Qy   439 ValGlyLeuValHisSerAsnIleAlaThrGluHisIleProSerProAlaLysVal 458
         ::|||||||||||||||||||||||||||||||||||||||||||
Db  1563 ATGGGGCTCGTGCATAGTAACACAGCACAGAGCATAATTCCAGCCCCTGCTAAGAAGTG 1622

Qy   459 ProArgLeuProAlaThrAlaAlaGluProGluAlaIleAlaIleSerAsnGlyGluHis 478
         |:|||||||||||||||||||||||||||||||||||||||||||
Db  1623 CCAAGACTCCCTGCAAACCTCGGCCGAAATCTGAATCAGCTGTCATCAGTAACGGGGAGCAC 1682
```

```
RESULT 3
AK031297
LOCUS       AK031297               3466 bp    mRNA    linear   HTC 18-SEP-2003
DEFINITION  Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
            enriched library, clone:6030403N11 product:pre B-cell leukemia
            transcription factor 1, full insert sequence.
ACCESSION   AK031297
VERSION     AK031297.1  GI:26327188
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99272253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   2053091
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3466)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3466
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6030403N11"
                     /db_xref="MGI:2394508"
                     /db_xref="taxon:10090"
                     /clone="6030403N11"
                     /sex="male"
                     /tissue_type="testis"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
     CDS             122..1684
                     /note="unnamed protein product; pre B-cell leukemia
                     transcription factor 1 [MGD|MGI.97495, GB|NM_008783,
                     evidence: BLASTN, 99%, match=124)
                     putative"
                     /codon_start=1
                     /protein_id="BAC27338.1"
                     /db_xref="GI:26327189"
                     /translation="MPGSLPLNABACWPKDVGIVALEIYFPSQYVDQAELEKVDGVDA
                     GKYTIGLGQARWGFCTDREDINSLCLTVVQKLMERHSLSYDCIGRLEVGTETIIDKSK
                     SVKSNLMQLFEESGNTDIEGIDTTNACYGGTAAVFNAVNWVESSSWDGRYALVVAGDI
                     AIYATGNARPTGVGAVALLIGPNAPLIFDRGLRGTHMQHAYDFYKPDMLSEYPVVDG
                     KLSIQCYLSALDRCYSVYRKKIRAQWQKEGKDKDFTLNDFGFMIFHSPYCKLVQKSLA
                     RMFLNDFLNDQNRDKNSIYSGLEARGDVKLEDTYFDRDVEKAFMKASSELPNQKTKAS
                     LLVSNQNGNMYYSSSVYSLASVLAQYSPQQLAGKRVGVFSYGSGLAATLYSLKVTODA
                     TPGSALDKITASLCDLKSRLDSRTCVAPDVFAENMKLREDTHHLANYIPQCSIDSLFE
                     GTWYLVRVDEKHRRTYARRPFTNDHSLDEGMGLVHSNTATEHIPSPAKVPRLPATSA
                     ESESAVISNGEH"
     polyA_signal    3447..3452
                     /note="putative"
     polyA_site      3466
                     /note="putative"
ORIGIN
```

```
Alignment Scores:
Pred. No.:             3
Score:                 3.77e-258     Length:        3466
                       2348.00       Matches:       452
Percent Similarity:    89.42%        Conservative:  13
Best Local Similarity: 86.92%        Mismatches:    13
Query Match:           93.51%        Indels:        42
                       11            Gaps:          1
DB:

US-10-622-516-2 (1-478) x AK031297 (1-3466)

Qy    1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    122  ATGCCTGGGTCACTTCCTTTGAATGCAGAGGGCTTGCTGGCCAAAAGATGTGGGAATCGTT  181
Qy     21  AlaLeuGluIleTyrPheProSerGlnThrValAspGlnAlaLeuGluLeuGluLysTyrAsp   40
Db         GCCCTTGAAAATCTACTTTCCTTCTCAAATAATGTCGATCAAGCTGAGTTGGAAAAATACGAT
Db    182  .............................................................  241

Qy     41  GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr   60
Db    242  GGTGTAGATGCTGGGAAAGTATACCATCGGGCCTTGGGCCAGGCATGGGGCTTCTGCACG   301

Qy     61  AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
Db    302  GATCGTGAAGACATCAACTCTTTGTGCATGACTGTGGTTCAAAACTGATGGAGAGACAT   361

Qy     81  AsnLeuSerTyrArgCysIleGlyTyrGlyAspThrGlyThrIleLeuIleLysAsp   100
Db    362  AGCCTTTCCTATAGATTGCATTGGGGGGCTTAGAGAGTTGGAACAGAACAATCATCGACAAA   421

Qy    101  SerLysLysSerValLysThrAsnLeuMetGlnLeuPheGluGlyAsnThrAspIle   120
Db    422  TCGAAATCACGTAAAGTTCTAAATTTGCATGCAGCTTTGATGGAGTCTCGGGAATAACAGATATA   481

Qy    121  GluGlyIleLeuAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValAla   140
Db    482  GAAGGAATAATAGATACAACAAATGCATGCTATGGGGGAACAGCTGCAGTCTTCAAATGCCCTG   541

Qy    141  AsnTrpIleGluSerSerSerTrpAsp                               149
Db    542  AACTGGGTCGAAATCCAGCTCTTGGGATGGACGATATGCCTCGGTAGTTGCAGGAGACATT   601

Qy    149  ----------------------------------------                  149
Db    602  GCTATATATGCCACAGGAAAATGCCAGACCTACAGGTGGAGCTGGAGCTGGCCCTGCTA   661

Qy    150  GlyLeuArgGlyThrHisMetGlnHis  158
Db    662  AATTGGGGCCAAAACGCTCCTCTAATTTTGACCGAGGGCTCCTGGGACACACATGACGAT   721

Qy    159  AlaTyrAspPheTyrGlyTyrProAspMetLeuSerGlnThrTyrProIleValAspGlyLeu   178
Db    722  GCCTATGACTTTTACAAGACCCTGACATGCTCTCCGAGTACCCTGTGGTCGACGGAAAGTTC   781

Qy    179  SerIleGlnCysTyrLeuSerArgAlaLeuAspAspArgCysTyrSerArgValTyrCysLysLysIle   198
Db    782  TCCATACAGTGCTACCTCAGCGCCCTGGACGCCCGCTGCTATTCGTCTACCGGAAAAGATC   841

Qy    199  HisGlnAlaGlnIleTrpGlnLysGlyAspAsnAspAspSerPheGlyPhe   218
Db    842  CGTGCCCAGTGCCAGAAAGGCAGAAAGGGATAAAGATTTTACCCTGAAATGAATTTGGCTTC   901

Qy    219  MetIleGlnPheHisSerProTyrCysGlyLysSerLeuValGlnGlyLysSerLeuLeuArgMetLeuLeu   238
Db    902  ATGAATCTTTCACTCACCATATTGTAAACTGGGTGCAGAAAATCTCTAGCTCGGAGTTCCCTG   961

Qy    239  AsnAspPheLeuAsnAsnArgAsnAlaAspAspLysAsnGlySerLeuLeuTyrSerGlyLeuGluAla   258
Db    962  AATGACTTTCTTAATGATCAAAACAGAGATAAAACAGTATTTACAGTGGTGACTGGAACC   1021

Qy    259  PheGlyAspValLysGluGluLysTyrThrTyrPheAspArgAspValGlyLysSerAlaPheMet   278
Db   1022  TTTGGGGACGTTAAATTAGAAGATAACTTACTTTGACAAGAGCGGGCATTTATG   1081

Qy    279  LysSerAlaSerSerGluLeuPheSerGlnIleLysThrLysSerAlaSerLeuLeuValSerArgAsnGln   298
Db   1082  AAGGCTAGTTCTGAGCTATTCAACCAGAAAAACAAAGGGGTCTTTGTCTTGTGTCTAATCAG   1141

Qy    299  AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuValAlaSerValAlaLeuAlaGlnIleTyr   318
Db   1142  AATGGAAATATGTACACATCCTCTGTCTATGGTTCTCTGGTGCCTTCTGTGGCTCTGGCACAGATTAC   1201

Qy    319  SerProGlnLnLeuUalAlaGlyLylysArgIleGluGlyValPheSerTyrGlySerGlyLeuAla   338
Db   1202  TCACTTCAGCAGTTGGCAGGGAAAGAAGGGGTTGGAGTGTTCTCTTACGGTTCGGGCTTGGCT   1261
```

Right column alignment (AY416670):

```
Db    339  AlaThrIleLeuTyrSerLeuLysValThrGlnAspAlaThrProGlySerAlaLeuLeuAspLys   358
Db   1262  GCCACCGCTGTACTCCCTTTAAAGTCACAGATGCCACCAGGAATCTGCCCTTGATAAA   1321

Db    359  IleThrArgAlaSerLeuCysAspSerLeuLysSerArgLeuAspSerArgAsnThrGlyValAlaPro   378
Db   1322  ATAACAGCAAGTTTATGTGACCTTAAATCAAGGCTTGACTCGAGAACTTGTGTGGCACG   1381

Db    379  AspValPheAlaGluAlaLysMetLeuLeuArgGluAspAspThrHisHisLeuValAsnTyrIle   398
Db   1382  GATGTCTTTGCTGAAAACATGAAAGCCTCAGAGGAGGACACACAATCACTTAGCCAACTATATT   1441

Db    399  ProGlnGlySerIleLeuAspPheLeuGlyThrTrpTyrLeuValArgValAspGlu   418
Db   1442  CCCCAGTTCAATAGACTCATCTTTGAAGGACATGGGTATCTGGTCAGAGTGGAATGAA   1501

Db    419  LysHisSerArgArgThrTyrAlaArgProThrProAsnAspAspAspThrLeuAspGlyGly   438
Db   1502  AAACACAGAAGGACTTACGCCCGGCCCTTCACAAATGACCACAGTTTGGATGAAGGA   1561

Db    439  ValGlyValHisSerAsnIleAlaThrGluHisIleProSerProAlaLeuLysSerVal   458
Db   1562  ATGGGGCTCGTCCATAGTAACACACAGCAACAGGCATATTCCAAGCCCCTGCTAAGAAAGTG   1621

Db    459  ProArgLeuProAlaThrAlaAlaGluProGluLuAlaAlaVaIIleSerGluAsnGlyGlyHis   478
Db   1622  CCAAGACTCCCTGCAACCTGCGGCCGAATCTGAAATCAGTAACGGGGGAGGCGAC   1681
```

RESULT 4
AY416670
AY416670

```
LOCUS       AY416670        1527 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Homo sapiens HMGCS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY416670
VERSION     AY416670.1  GI:39772630
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1527)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 1527)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..1527
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene            <1..>1527
                     /gene="HMGCS2"
                     /locus_tag="HCM5968"
ORIGIN
```

```
Alignment Scores:
Pred. No.:
Score:          1.36e-162    Length:   1527
                1513.00      Matches:  282
```

```
US-10-622-516-2 (1-478) x AY416670 (1-1527)

Percent Similarity:   75.16%      Conservative:   69
Best Local Similarity: 60.39%     Mismatches:     74
Query Match:           60.25%     Indels:         42
DB:                    29         Gaps:           1
```

```
Qy     4 SerLeuProLeuAsnAlaGlyLeuAlaCysTrpProLysAspValGlyIleValAlaLeuGlu 23
           :::|||:::|||||||:::|||||||||:::|||||:||::::|||::::|:::|
Db   121 GCTGTCCCCCTGGCCAAAACAGATACTTGGGCAAAAGACGTGGGGCATCCTGGCCCTGGAG 180

Qy    24 IleTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAsp 43
           :::|||||||||:::||||||:::||||||||||||:::|||::||||::::
Db   181 GTCTACTTCCCAGCCCAATATGTGGACCAAGCTGAGAAGAAGTATAACAATGTGGAA 240

Qy    44 AlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrAspArgGlu 63
           |||:::||||||||||||:||||||:::|||||||||||||:::|||:::|||
Db   241 GCAGGAAAGTATACAAGTGGGCTTGGGCAGGCCAAAATGGGCTTCTGCACAGATCGAGAG 300

Qy    64 AspIleAlaAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsnAsnLeuSer 83
           :::|||:::||||||||||||:||||||||||:::||||||||||:::|||:::|||
Db   301 GACATCAACTCTCCTGTGCCCTGACGGTGGTGCAACGGGCTGATGGAGCGTCCCA 360

Qy    84 TyrAspCysIleGlyTyrArgLeuGluValGlyValGlyThrGlyThrLeuIleGluAspProLysLysSer 103
           ::|||:::||||||||||||:::|||||||||||||:::|||||||||||:::
Db   361 TGGGACTGCATCGGGTACCGGCTGGAAGTAGGCACTGGAACCATCATTGAGGACCCAAAGT 420

Qy   104 ValLysThrAsnLeuMetGlnIleLeuPheGluGluGlyIleIle 123
           |||::::|:::||||||||:::||||||:::|||||||||||
Db   421 GTCAAAACAGTGCTCATGGAACTCTTCCACGGAATGCCAACTACTGATATTGAGGGCATA 480

Qy   124 AspThrThrArgAsnAlaCysTyrArgGlyTyrGlyThrArgAlaLeuAlaValPheAsnAlaValAlaAsnTrpIle 143
           :::|||:::|||||||:::||||||||||||:::|||||||||||||:::
Db   481 GATACCACCAATGCCTGCTACCGTGGGTACTGCCCTTCAAATGCTGCCAACTGGATG 540

Qy   144 GluSerSerSerTrpAsp--- 149
           -------------------
Db   541 GAGTCCAGTTCCTGGGATGGTTCGTTATGCCATAGGTGGTGTGTGGAGACATTGCCGTCTAT 600

Qy   149 -------- 149
           --------
Db   601 CCCAGTGGTAATCGTCCCACAGGTGGGGCCGGAGCTGTGGGCTATGCTGAATGGGGCC 660

Qy   150 ---GlyLeuArgGlyThrHisMetGlnHisAlaThrYrAsp 161
           ::::||||||||||:::||||||:::||:::
Db   661 AAGGCCCCTCTGCCCCTGGAGCGCAGGGGCTGAGGGAACCCATATGGAGAAATGTGTATGAC 720

Qy   162 PheThrArgLysProAspMetLeuSerGluTyrProIleValAlaSerProGlyLysLeuSerIleGln 181
           ||||||||||||:::|||:::|||:::|||||||||||||
Db   721 TTCTACAAAACCAAATTGGCCTCGGAGTACCCAATAGTGGCAAGCCCTTCCATCCAG 780

Qy   182 CysTyrLeuSerLeuAlaLeuAspArgCysTyrSerTyrSerValTyrCysGlyLysGlyIleHisAlaGln 201
           :::||||||||:::|||:::||||||:::|||||||||:::|||||||||
Db   781 TGCTACTTGCGGGGCCTTGGATCGGAGTTACACATCAACCGTAAAAAAAATTCCAGAATCAG 840

Qy   202 TrpGlnIleLysGlyLysAspAsnAspPheThrLeuAsnAspAspPheGlyPheMetIlePhe 221
           |||:::|||||||:::|||||||||||||:::||||||||
Db   841 TGGAAGCAAAGCTGGCAGGGATCAACACCCTTCACCCTTGACGATTTACAGTACATGATCTTT 900

Qy   222 HisSerProTyrCysGlyLysValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPhe 241
           |||:::|||||||:::|||||||:::||||||||:::||||||||
Db   901 CATACACCCTTTGCAAGTAGGTTCAGAAGTCCTGGCTCGCATGTTCAATGACTTC 960

Qy   242 LeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluLeuAlaPheGlyAsp 261
           ||||||:::|||||||:::|||:::||||||||:::|||||||||
Db   961 CTGTCAGCCAGCAGTGACAACAAACCAGCTTATATAAGCGGGCTGGAGCTGGCCTTTGGGGG 1020

Qy   262 ValLysLeuGluAspThrThrYrPheAspArgAspValGlyLysAlaPheMetLysAlaSer 281
           |||:::|||||||:::|||:::|||||||:::|||||||:::|||||||||||||
Db  1021 CTAAAGCTGGAAGACACCTACACCAACAAGGACCTTCTAAAGGCCCTCT 1080

Qy   282 SerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsn 301
           :::||||||||:::||||||
Db  1081 CAGGACCATGTTCGACAGAAGAAAACCAAAGGCTTCCCTTTACCTCCACAACAATGGGAAC 1140

Qy   302 MetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnThrYrSerProGln 321
           |||:::|||||||||:::|||::::|||||||||||||:::|||:::
Db  1141 ATGTACACCTCATCCGTACGGGGTCGGGTGCCTGGCCTTCTGTGTCCCACACTCTGCCCAA 1200

Qy   322 GlnLeuAlaGlyIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeu 341
           |||:::||||||:::||||||:::||||||||:::
Db  1201 GAACTGGCTGGCATCGGAGTGTTCTCCTATGGCTCGGGTTTAGCAGCAAGTTC 1260

Qy   342 TyrSerLeuLysValThrGlnAsnAlaThrProGlySerAlaSerLeuIleThrAla 361
           |||:::||||||:::||||||:::||||||||:::||||||
Db  1261 TTTTCATTCGAGTATCCCAGGATGCTGCTTCCAGGCGCTCCCCCCTGGCACCAGTGGTGTCC 1320

Qy   402 SerIleAspSerLeuPheGluGlyTyrThrTrpTyrLeuValArgValGlyLysHisArg 421
           :::|||:::||||||||:::||||||:::||||||||||||:::|||
Db  1381 ACAGAAAATAATGAACCAAAGAGAGCAATTCTACCATAAGGTGAAATTCTCCCCACCTGGT 1440

Qy   422 ArgThrTyrAlaArgArgPro 428
           ||||||:::|||||||:::
Db  1501 CGAAAGTATCCCGGCGTCC 1521
```

```
RESULT 5
AY416672
LOCUS       AY416672               1527 bp    DNA     linear   GSS 17-DEC-2003
DEFINITION  Mus musculus HMGCS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY416672
VERSION     AY416672.1  GI:39772632
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1527)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2  (bases 1 to 1527)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT     This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
     source          1..1527
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
     gene            <1..>1527
                     /gene="HMGCS2"
                     /locus_tag="HCM5968"
ORIGIN
Alignment Scores:
Pred. No.:            2.74e-160          Length:       1527
```

```
Score:                    1493.00    Matches:         276
Percent Similarity:       75.27%     Conservative:     74
Best Local Similarity:    59.35%     Mismatches:       73
Query Match:              59.46%     Indels:           42
DB:                       29         Gaps:              1

US-10-622-516-2 (1-478) x AY416672 (1-1527)

Qy      6 ProLeuAsnAlaGluAlaCysThrProProLysAspValGlyIleValAlaLeuGlyIleTyr 25
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127 CCCCTGGCCAAAACAGATACATGGCCAAAAAGATGTGGGCATCCTTGGCCCTGGGAGGTCTAT 186

Qy     26 PheProSerGlnTyrValAlaspGlnLeuGlyGlyTyrAspGlyValAspAlaGly 45
          ||||||||||||||||||||:||||||||||||||||||||||||||
Db    187 TTTCCAGCCCAAATATGTGGACAACAACTGACCTGGAAAAGTTCAACAAATGTGGAAGCAGGG 246

Qy     46 LysTyrThrIleGlyLeuGlyIleAlaLysMetGlyPheCysThrThrAspArgGluAspIle 65
          |||:||||:||||||||||||||||||||||||||||||||||:||||||||||
Db    247 AAGTATACACAGTGGGCTTGGGCCAAGCCCGTATGGGCTTCTGTCAGTCCAGGAGGACATC 306

Qy     66 AsnSerLeuCysMetThrValValGlnArgAsnAsnLeuSerTyrAsp 85
          |||||||||||||:|||||||||||||||||||:::||
Db    307 AACTCCCTGTGCACCGTGGTTCAGAGGAACCGACCAAAAGCTGCCGTGGGAT 366

Qy    307 AACTCCCTGTGCACCGTGGTTCAGAGGAACCGACCAAAAGCTGCCGTGGGAT 366
```

RESULT 6
LOCUS   AK004865   3287 bp   mRNA   linear   HTC 20-SEP-2003
DEFINITION   Mus musculus adult male liver cDNA, RIKEN full-length enriched
               library, clone:1300002P16
               product:3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2, full
               insert sequence.
ACCESSION   AK004865
VERSION   AK004865.1   GI:12836370
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
               Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
               prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
               Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
               Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
               Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
               Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
               Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
               Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
               sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4

```
Db   1087 ATGTTCAACCAGAGAGACCAAGGCCTCCCTTCACCTCTCACAAAATAATGGGAACACATGTAC 1146

Qy    324 AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSer 343
          |||||||||||||||||||||||||||||||||||||:::|||
Db   1207 GCTGGCCTTCCAGGATTGGGAGCCCTTCTCCTACGGCTCAGGCTTAGCAGCAGAGTTCTTTCA 1266

Qy    344 LeuLysValThrGlnAspAlaThrProGlySerAlaLeuAspLysIleThrAlaSerLeu 363
          |||||||||||||||||||||:||||||||||||||||||||:|||||
Db   1267 TTCCGAGTGTCCAAGGATGCTTCCCCAGGTTCCCCCTCGGAGAAACTGGTGTCTAGTGTG 1326

Qy    364 CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAlaValPheAlaGlu 383
          |||||||||:||||||||||||||||||||||||||||:::|||
Db   1327 TCAGATCTGCCCAAACGTCTAGACTCCCGGACAGGCATGTCCCCTGGAGGAATTCACAGAA 1386

Qy    384 AsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGlnGlySerIle 403
          |||||||||||||:||||||||||||||||:|||
Db   1387 ATAATGAACCAGAGAGACCAAGGCAATTTTTACCACAAGGTGAACTTTCTCCACCTGGACACA 1446

Qy    404 AspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlyLysHisArgThr 423
          |||:||||||||||||||||||||||||||||||||||
Db   1447 AGCAATCTTCTTCCACCAGGTACTTGGTACCTTGAACGAGTGGATGACATCGCAGGAAG 1506

Qy    424 TyrAlaArgArgPro 428
          ::||||||||
Db   1507 TATGCCCGGCGTCC 1521
```

```
AUTHORS     The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3287)
AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,Y., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            This clone has been submitted in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGACGGGCCCCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence[5'
            GAGAGAGAGACGGGCCCCAAAATATTAATTAATAAACCCCCCCCC 3'], cDNA was
            cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
            XhoI. Host: SOUR.
FEATURES             Location/Qualifiers
     source          1..3287
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1300002P16"
                     /db_xref="MGI:1907269"
                     /clones="1300002P16"
                     /clones="taxon:10090"
                     /sex="male"
                     /tissue_type="liver"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             52..1578
                     /note="unnamed protein product;
                     3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2
                     (MGD|MGI:101939)
                     putative"
                     /codon_start=1
                     /protein_id="BAB23626.1"
                     /db_xref="GI:12863371"
                     /translation="MQRLLAPARRVLQVKRAMQETSLTPAHLLSAAQQRFSTIPAPL
                     AKTDWPKDVGILALEVYFPAQYVDQDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI
                     NSLCLTVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKTVLMELFQDSGNTDIEGI
                     DTTNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNARPTGGAGAVAMLI
                     GPKAPLVLEQGLRGTHMENAYDFYKPNIASEYPLVDGKLSICYLRALDRCYAAYRKK
                     IQNWKQAGNNQPFTLDDVQYMIFHTPFCKMVQKSLARLMFNDFLSSSSSDKQNNLYKG
                     LEAFRGLKLEETYTNKDVDKALLKASLDMFNQKTKASLYLSTNNGNMYTSSLYGCLAS"
ORIGIN
            polyA_signal    3266..3271
                            /note="putative"
            polyA_site      3287
                            /note="putative"
```

US-10-622-516-2 (1-478) x AK004865 (1-3287)

```
Alignment Scores:
Pred. No.:              9.846e-160
Score:                  1493.00
Percent Similarity:     75.27%          Length:        3287
Best Local Similarity:  59.35%          Matches:        276
Query Match:            59.46%          Conservative:    74
DB:                     11              Mismatches:      73
                                        Indels:          42
                                        Gaps:             1
```

```
Qy    1  LLSHHSAQELAGSRIGAPSYGGSGLAASFSFRVSKDASPGSPLPEKLVSSVSDLPKRLD      60
Db       SRRRMSPEEFTEIMNQREQFYHKVNFSPPGDTSNLFPGTWYLERVDEMRRKYARCPV

Qy    6  ProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleValAlaLeuGluValTyr    25
         |||::|.:|||:::||:|  |||||||||||:|  |||:||:.||||:||||:|||||||
Db  178  CCCTGGCCAAAACAGATACATGGCCAAAGATGTGGGCCATCCTTGCCCTGGAGGTCTAT   237

Qy   26  PheProSerGlnTyrValAspProAspLeuGluLysPheAsnAsnValGluAlaGly      45
         ||:|  ::|||:|||.|:|:|||||||||||:|||||||||:|||||||||:||:||
Db  238  TTTCCAGCCCAATATGTGGACCCAGATCTTGAGAAGTTCAACAATGTGGAAGCAGGG    297

Qy   46  LysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysSerThrAspArgGluIle    65
         :::||||||:||||.||||||||||:||||||||:||:||:|:.|||||.|||||:|||
Db  298  AAGTATACAGTGGGCCTTGGGCAGACCCGTATGGGCTTCTGCAGTCCAGGACATC    357

Qy   66  AsnSerLeuCysMetThrValGlnArgLeuMetGluArgAsnAsnLeuSerTyrAsp      85
         ||||||||:|:|||:|||:||||||||:||||:||||:|||:||.||.|||:|||:
Db  358  AACTCCCTGTGCCTGACCAGTGGTACACAGAGGCTGATGGAGAGGACTCTCCGTGGGAT    417

Qy   86  CysIleGlyArgLeuGluValGlyThrGluThrIleIleIleAspLysSerLysValLys   105
         |||:||||||:||||||||||:||||:||:||:|||::|||||||||:||:|||:|||||
Db  418  TGCATTGGGCGGCTCGAAGTGGGCACCGAGACCATCATTGACAAGTCCAAGGCTGTCAAA   477

Qy  106  ThrAsnLeuMetGluLeuPheGluGluSerGlyAsnThrAspIleGluGlyIleAspThr   125
         |||:|||||::||:||||||||||:||||:||:|||:|:|||||||:||.:|:|:|||||
Db  478  ACAGTCGCTCATGGAACTGTTCCAGGATTCAGGGCATAGATACC    537

Qy  126  ThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaAlaAsnTrpIleGluSer   145
         ||:|||:||||:||||||:||||:|||.|||||||||:|:||||||:|||:|||:|||:|
Db  538  ACCAACGCCTGTTATGGCGGCACACCCTCTTCAATGCTGCCAACTGGATTGAGTCC    597

Qy  146  SerSerTrpAsp------------------------------                    149
         |||:|||:|
Db  598  AGCTACTGGGATAGGTCGGTATGCCCTGGTGTTCTGTGGGGTGACATTGCCAGGTCTAT   657

Qy  149  ------------------------------------------                   149
Db  658  GGTAACGCCCGCCCCACAGGTGGTGCTGGGCAAGTGCAATGCTGATCGGGCCAAGGCC   717

Qy  150  ------------------GlyLeuArgGlyThrHisMetGlnAsnAlaTyrAspPheTyr   163
                          ||||||||:||||:|||:|:|||.|||:|:|||||||||||
Db  718  CCTCTGGGTCCTCGACGACAAGGGCTGAGGGGAACTCACATGGAGAACGCGTACGACTTCTAC   777

Qy  164  LysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlyCysTyr   183
         :||||:||:||.||||:|||:|||:||||||||||:||:|||||:|:||:||||||||:
Db  778  AAACCAAACTTGGCCTCAGATATGCTATGGCCCTGTGGATGGTGACCTGTCTATCCAGTCTAC   837

Qy  184  LeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAlaGlnInTrpGln   203
         |||||:|||:|||||:||||||:::||:|:||:::||:||||:|:|||:|:||:|:||
Db  838  CTGCGGGCCCTTGGACTGGATCGATGCTATGCCCGGCAGAAGATCCAGAAATGGCAGTGGAAG   897

Qy  204  LysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIlePheHisSer   223
         :::|  :|||.||::|::|:||.:|  ||.|||.|||||||||.||:|||:||:::||:
Db  898  CAAGCTGGAAACAACCAGCCTTTCACCCTCGATGATGTGCAGTATATGATCTTTCACA   957

Qy  224  ProTyrCysLysLeuValGlnLysSerLeuAlaArgMetLeuAsnAspPheLeuAsn     243
         |||.||:||::|||||:|:||::|||::||:|:||.::|||:||.|||||:::
Db          ||||::||::
```

```
Db    958  CCCTTTTGCAAGATGGTCCAGAAATCCCTGGCTCGGTTGATGTTGTTCAATGACTTCCTGTCA
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    244  AspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLys  263

Db   1018  TCCAGCAGTGACAAACAGAAACAACTTATACAAGGACCCTTCAGGGGTCTAAAG  1077
                ||||||||||||||||||||||||||||||||||||||||||||
Qy    264  LeuGluAspThrTyrPheAspArgAspValGluLysAlaPheMetLysAlaSerSerGlu  283

Db   1078  CTGGAAGAGACATACTACGACAGACAAGGATGTAGACAAGGCTCTTTGAAGGCCTCCTCGAAC 1137
                ||||||||||||||||||||||||||||||  :::::
Qy    284  LeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetTyr  303

Db   1138  ATGTTCAACCAGAGAGACACCAAGGCCCTCCCTTTACCTCTCCACAAATAATGGGAACATGTAC 1197
                ||:::||||||||||||||||||||||||||::||||||||||
Qy    304  ThrSerSerValTyrGlySerLeuValAlaSerValLeuAlaGlnTyrSerProGlnGluLeu  323

Db   1198  ACCTCTTCCCTCTATGGCTCGCTTGGCCTTCTCACTTCTCTCACCCACTCTGCCCAAGAAATTG 1257
                |||||||||||||||||||||||||||||||||
Qy    324  AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuLeuAlaAlaThrLeuTyrSer  343

Db   1258  GCTGGCTTCCAGGATTGGGACGGCTTCTCCTTACGGCCTCAGGCTTAGCCAAGTTTCTTTTCA 1317
                |||||:::|||||||||||||||||||||::|||:::
Qy    344  LeuLysValThrGlnAspAlaThrProGlyLysSerAlaLeuLeuAlaThrAlaSerLeu  363

Db   1318  TTCCAGGTGTCCAAGGATGCTTCCCCCAGTTCCCCCTGGAGAAACTGGTGTTCCTCTTCGTGTG 1377
                ||:::|||||||||||||||||||||||||||||:::||
Qy    364  CysAspLeuLysSerArgLeuAspSerArgThrGlyValAlaProAspValPheAlaGlu  383

Db   1378  TCAGATCTTGCCCAAACGTTCTAGACTCTCCCGGACGACGGCATGGGAATTCACAGAA 1437
                ||||||||||||||::::|:::|||||||::||||
Qy    384  AsnMetLysLeuArgGluAspThrHisHisLeuValAsnTyrIleProGlnGlySerIle  403

Db   1438  ATAATGAACCAGAGAGGAGCAATTTACCACAAGGTGAACTTCTCTCCACCTGGAGACACA 1497
                |||||:::|||||||||||||||||||::|||||
Qy    404  AspSerLeuPheGluGlyThrTrpTyrLeuValValArgValAspGlyHisArgArgThr  423

Db   1498  ACAAATCCTTCCCAGGTACTTGGTACCTTGGACGATGGAGATGCATCCCAGGAAG 1557
                |||||||||||||||||||||||||||
Qy    424  TyrAlaAlaArgGlyPro  428

Db   1558  TATGCCCGGGTGTCCC 1572
                |||||
```

```
                   putative"
                   /codon_start=1
                   /proteín_id="BAB23657.1"
                   /db_xref="GI:12836439"
                   /translation="MQRLLAPARRVLQVKRAMQETSLTPAHLLSAAQQRPSTIPPAPL
                   AKTDTWPKDVGILALEVYFPAQYVDQTDLEKFNNVEAGKYTVGLGQTRMGFCSVQEDI
                   NSLCLTVVQRLMERTKLPWDAVGRLEVGTETIIDKSKAVKVIMELFQDSGNTDIEGI
                   DTTNACYGGTASLFNAANWMESSYWDGRYALVVCGDIAVYPSGNARPTGGAGAVAMLI
                   GPKAPLVLEQGLRGTHMENAVDFYKPNLASEYPVLDGKLSIQCYLRALDRCYAAYRKK
                   IQNQWKQAGNQPFTLDDVQYMIFHTPFCKMVQKSLARLMENDFLSSSSDKQNNLYKG
                   LEAFRGLKLEETYTNKDVDKALLKASLDMFNQKTKASLLYLSTNNGNMYTSSLYGCLAS
                   LISHHSAQELAGSRIGAFSYGSGLAASFFSFRVSKDASPGSPLEKLVSSVSSDLPKRLD
                   SRRRMSPEEFTEIMNQREQFYHKVNFSPPGDTSNLFPGTWLERVDEMHRRKYARCPV
                   "
     polya_signal   2503..2508
                   /note="putative"
     polya_site     2524
                   /note="putative"
ORIGIN
```

```
Alignment scores:
Pred. No.:
Score:                       1.08e-159
                             1491.00
Percent Similarity:          75.27%
Percent Identity:            75.14%
Best Local Similarity:       59.14%
Query Match:                 59.38%
DB:                          11

Length:          2524
Matches:         275
Conservative:    75
Mismatches:      73
Indels:          42
Gaps:            1
```

US-10-622-516-2 (1-478) x AK004902 (1-2524)

```
Qy    6  ProLeuAlaGluAlaAlaCysThrProProLysAspValAlaGlyIleValIleTyr   25
            |||  |||||       |  :::  ||||||||||||||||::||||||||||
Db  176  CCCCTGGCCAAAACAGATACATGGCCACAAAAGATGGGCATCGTTCCCCTGGAGGTCTAT  235

Qy   26  PheProSerGlnThrTyrValAspGlnAlaGluLeuGluLysTyrAspGlyValAspAlaGly   45
            :::||||||     ::|||||||||:::||||||||||||||||||||||:::||||||
Db  236  TTTCCAGCCCAAATATGTGGACCAAGCAGAAAACTGAAAAGTTCAACAATGTGGAAGCAGGG  295

Qy   46  LysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThrArgArgGluAspIle   65
            |||||||||||||||||||||:::||||||||||||:::|||:::||||||||||||||
Db  296  AAGTATACAGTGGGCCTTGGGCCAGACCCGTATGGGTTCTGTTCAGTCCAGGAGGACATC   355

Qy   66  AsnSerLeuCysMetThrValValGlnArgAsnAsnLeuSerTyrTyrAsp   85
            |||:::|||||||||||||||||||||||||||||:::|||:::||||||
Db  356  AACTCCCTGTGCCTGACAGTGGTACAGAGGAACAGCACAAAAGCTGCCTGGGAT   415

Qy   86  CysIleGlyArgLeuGluValGlyThrGluThrIleIleAspLysSerLysValLys  105
            |||:::|||||||||||||||||||||||||||||:::||||||||||:::||||||
Db  416  GCTGTGGGCCGTCTGGAAGTGGGCACCGAGACCATCATTGACAAGTCCAAGGTGTCAAA   475

Qy  106  ThrArgLeuMetGlnLeuProPheGluGluSerGlyAsnThrAspIleGluGlyIleIleAspThr  125
            ||||||||||||:::|||||||||:::|||||||||||||||||||:::
Db  476  ACAGTGCTCATGGAACTGTTCCAGGATTCAGGGAATACGGACATAGAATACC   535

Qy  126  ThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaAlaAsnTrpIleGluSer  145
            |||::||||||||||||||||||||:::||||||||||||||||||||||||||:::
Db  536  ACCAACGCCTGTTATGGCGGACACCTCCCTCTTCAATGCCGCCAACTGGATGGAGTCC   595

Qy  146  SerSerTrpAsp-----------------------------------------   149
            ||  |||||||
Db  596  AGCTACTGGGATGTCGGCTATGCCCTGGTCGTGTGTGGCGACATTGCAGTCTACCCGAGT  655

Qy  149  --------------------------------------------------   149
Db  656  GGTAACGCCCGCCCCACAGGTGGCGTGGCAGTGGCCATGCTCATCGGGCCCAAGGCC   715

Qy  150  -------------GlyLeuArgGlyThrHisMetGluAsnAlaValAspPheTyr   163
                        |||:::||||||   ||||||  :::|||
Db  716  CCCTCTGGTTCCTGAGCAAGGGCTACAGGGAACACGGTACGACTTCTAC   775

Qy  164  LysProAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyr  183
            |||  ||:::||    |||:::|||:::
Db  776  AAACCAAACTTGGCCTCAGAGTATCCACTGGTGGATGGGAAGCTGTCTATCCAGTGCTAC  835
```

```
Qy  184  LeuSerAlaLeuAspArgCysTyrAlaAlaTyrArgLysLysIleGlnAsnGlnTrpGln  203
            ||| |||||||||||:::||||||||||:::|||||||||||||||||||:::
Db  836  CTGGCAGCGCTGGACCGATGCTATGCAGCCTACCGCAAGAAGATCCAGAATCAGTGGAAG  895

Qy  204  LysGlnGlyAsnGlnProPheThrLeuAspAspValGlnTyrMetIlePheHisThr  223
            |||:::||||||||||||||||||||:::||||||||||||||||:::||||||
Db  896  CAAGCTGGAAACAACCAGCCCTTCACCCTGGATGATGTGCAGTATATGATCTTCCACA  955

Qy  224  ProTyrCysLysMetValGlnLysSerLeuAlaArgMetLeuGluAsnAspPheLeuAsn  243
            |||:::||||||||||||:::||||||||||||||||:::||||||||||||||||:::
Db  956  CCCTTTTGCAAGATGGTCCAAAAATCCCTGGCTCGGTTGATGGTTCAATGACTTCCTGTCA  1015

Qy  244  AspGlnAsnAsnArgAspLysAsnSerIleIleTyrSerGlyLeuGluGlyAlaPheGlyAspLys  263
            |||:::|||||||||:::||||||||||||||||||||||||||||:::
Db  1016  TCCACGAGTGCACAAACAGAACAACTTATACAAGGGCCCTTCAGGGGTTAAAG  1075

Qy  264  LeuGluAspThrTyrPheAspArgAspValGlyLeuAlaPheMetLysAlaSerSerGlu  283
            ||||||:::||||||||||||||||:::||||||||||||||||||:::||||||
Db  1076  CTGGAAGAAACCTACACCAACAAGGATGTAGACAAGGCTCTTTTGAAGGCCTCCCTGGAC  1135

Qy  284  LeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGlyAsnMetTyrTyr  303
            |||:::||||||:::||||||||||||||||||||||:::|||||||||||||||
Db  1136  ATGTTCAACCAGAAGACCAAGGCCTCCCTTCTGGTCAGCAATGGGAACATGTAC  1195

Qy  304  ThrSerSerValTyrGlySerArgLeuAlaSerValLeuAlaGlnInTyrSerProGlnGluLeu  323
            |||||||||||||||||:::|||||||||||||||||||||||:::||||||
Db  1196  ACCTCTTCCGTCTACGGCTCCCGCCTAGCCAGTGTTCTGGCCCACTTAATCTCTGCCAAGAATTG  1255

Qy  324  AlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSer  343
            |||:::|||||||||:::|||||||||||||||:::
Db  1256  GCTGGCCTCCAGGATTGGAGCCTTCTCCTACGGCTCAGGTCTAGCTGCCACTTTCTTTCA  1315

Qy  344  LeuLysValThrArgGlnAsnAlaThrProGlySerAlaLeuLeuLeuThrAlaSerLeu  363
            |||:::||||||||||||||||:::||||||||||:::||||||||||||||:::
Db  1316  TTCCGAGTGTCCAAGCAGATGCTTCCCCCAGGTTCCCCACGGAAACTGGTGCTCTAGTG  1375

Qy  364  CysAspSerLeuLysSerArgLeuAlaSerValLeuAlaGlnIleValAlaProAlaValPheGluGlu  383
            |||:::|||||||||||||||:::||||||||||:::
Db  1376  TCAGATCTGCCCAAACGTCTAGCTCCCGGAGACCATCGTCCCCGAGGGAATTCACAGAA  1435

Qy  384  AsnMetLysLeuArgGluAspThrHisSerLeuValLeuAsnTyrIleLeuProGlySerIle  403
            ||||||||||||:::||||||||||||||||:::
Db  1436  ATAATGAACACGAGGAGGACAATTTTACACACACCGGGAATCTCTCCACCTGGCAGCATC  1495

Qy  404  AspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGluMetHisArgArgThr  423
            |||:::||||||||||||||||:::||||||||||||||||||:::
Db  1496  AGCAATCTCTTCCCAGGTACTTGGTATCTGGTGGATGAGATGCATCGGCGGAAG  1555

Qy  424  TyrAlaArgArgPro  428
            |||||||||
Db  1556  TATGCCCGGTGTCC  1570
```

```
RESULT 8
BX441322

LOCUS       BX441322               1083 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  CS0DF016YJ04 5-PRIME, mRNA sequence.  Homo sapiens cDNA clone
            CS0DF016YJ04 5-PRIME, mRNA sequence.
ACCESSION   BX441322
VERSION     BX441322.1  GI:30789948
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1083)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
  COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
```

Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3098.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF016E02QP1&cluster=3098.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF016E02QP1.

FEATURES   Location/Qualifiers
     source   1..1083
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF016Y204"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a Not1-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."

ORIGIN

US-10-622-516-2 (1-478) x BX441322 (1-1083)

Alignment Scores:
Pred. No.:
Score: 9.35e-148
Percent Similarity: 1382.00   Length: 1083
Best Local Similarity: 86.23%   Matches: 263
Query Match: 86.23%   Conservative: 0
DB: 55.04%   Mismatches: 0
                      13   Indels: 42
                                       Gaps: 1

Qy   1 MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 168 ATGCCTGGMTCACTTCCTTTGAAATGCAGAAGCTTGCTGGCCCAAAAAGATGTGGGAAATTGTT 227

Qy   21 AlaLeuGluIleTyrPheProSerGlnThrTyrValAspGlnAlaGluLeuGluLysTyrAsp   40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 228 GCCCTTGAGATCTATTTCCCTTCCCAAATATGTTGATCAAGCAGAGTTGGAAAAATATGAT 287

Qy   41 GlyValAspAlaGlyLysTyrThrIleGlyLeuGlyGlnAlaLysMetGlyPheCysThr   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 288 GGTGTAGACGCTGGAAAGTATACCATTGGCTTGGGCCAGGCCAAGATGGGCTTCTGCACA 347

Qy   61 AspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuMetGluArgAsn   80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 348 GATAGAGAGGATATAAACTCTCTTTGCATGACTGTGGTTCAGAATCTTATGGAGAGAAAT 407

Qy   81 AsnLeuSerTyrAspCysIleGlyArgLeuGluValAlaGlyThrGluThrThrIleLeuAspLys 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 408 AACCTTTCCTATGATTGCATTGGGCGGCTGGAAGTTGGAACAGAGACAAATCATCGACAAA 467

Qy 101 SerLysSerValLysThrAsnLeuMetGlnLeuPheGluGlyLeuGlyAsnThrArgAsn 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 468 TCAAAGTCTGTGAAGACTAATTTGATGCAGCTGTTTGAAGAGTCTGGGAATACAGATATA 527

Qy 121 GluGlyIleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaLeuPheAsnAlaVal 140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 528 GAAGGAATCGACACAACTAATGCATGTTATGGAGGCACAGCTGCTCTGTTTAATGCTGTT 587

Qy 141 AsnTrpIleGluSerSerSerTrpAsp------------------------------- 149
         ||||||||||||||||||||||||||||
Db 588 AACTGGATTGAGTCCAGCCTCTTGGGATGGACGGTATGCCCTGGGTAGTTGCACGAGATATT 647

Qy 149 --- 149
Db 648 GCTGTATATGCCACAGGAAAATGCTAGACCTACAGGTTGGAGCAGTAGCTCTGCTA 707

Qy 150 ------------------------GlyLeuArgGlyThrHisMetGlnHis 158
                                 |||||||||||||||||||||||||||
Db 708 ATTGGGCGCAAATGCTCCTTTAATTTTGAACCAGAGGCTTCGTTGGGACACATATGCCAACAT 767

RESULT 9
AY416671

LOCUS   AY416671   1420 bp   DNA   linear   GSS 17-DEC-2003
DEFINITION   Pan troglodytes HMGCS2 gene, VIRTUAL TRANSCRIPT, partial sequence,
             genomic survey sequence.
ACCESSION   AY416671
VERSION   AY416671.1 GI:39772631
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM   Pan troglodytes
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 1420)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
             gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2 (bases 1 to 1420)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
             Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
             Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
             Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
             Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
             them based on alignment.
FEATURES   Location/Qualifiers
     source   1..1420
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene   <1..>1420
                     /gene="HMGCS2"
                     /locus_tag="HCMS968"
ORIGIN

Alignment Scores:
Pred. No.:
Score: 2.17e-136
Percent Similarity: 1285.00   Length: 1420
Best Local Similarity: 72.55%   Matches: 243
Query Match: 58.00%   Conservative: 61
DB: 51.17%   Mismatches: 73
                      29   Indels: 42
                                       Gaps: 1

Qy 159 AlaTyrAspPheTyrLysProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 768 GCCTATGATTTTTACAGGCCTGATATGCTATCTGAATATCCTATAGTAGATGGGAAAACTC 827

Qy 179 SerIleGluCysTyrTyrLeuSerAlaLeuAspAlaArgCysTyrSerSerValTyrCysLysLysIle 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 828 TCCATACAGTGTTACTTCAGTGCATTAGACCGTTGCTACTCCTGTCTACTCGACAAAAGATC 887

Qy 199 HisAlaGlnThrTrpGlnLysGluGlyAsnAspLysAspPheThrLeuAsnAspPheGlyPhe 218
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 888 CATGCCCAGTGGCAGAAAGAGGGAAATGATAAAGATTTTACCTTGAATGATTTTGGCTTC 947

Qy 219 MetIlePheHisSerProThrTyrCysGlyLeuValGlnLeuValAlaArgMetLeuLeu 238
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 948 ATGATCTTTCACTCACCATATTGTAAACACGGTTCAGAAAATCTTCAGCTCGGATGTTGCTG 1007

Qy 239 AsnAspPheLeuAsnAspGlnAsnArgAspLysAsnAsnSerIleTyrSerGlyLeuGluAla 258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1008 AATGACTTCCTTAATGACCAGAATAGAGATAAAAATAGTATCTATAGTGGGCCTGGAAGCC 1067

Qy 259 PheGlyAspValLys 263
         |||||||||||||||
Db 1068 TTTGGGGATGTWAAA 1082

US-10-622-516-2 (1-478) x AY416671 (1-1420)

```
Qy     17  ValGlyIleValAlaLeuGluIleTyrPheProSerGlnTyrValAspGlnAlaGluLeu  36
               |||||:||::
Db    160  GTGGGCCATCCTGNNNCTGGAGGTCTACTTCCNNNNNNNNNNNNNNNNNNTGACCTG  219

Qy     37  GluLysTyrAspGlyValAspAlaGlyIleGlyTyrTyrIleGlyTyrLeuGlyGlnAlaLysMet  56
               :::||||
Db    220  GAGAAGTATAACAAATGTGGAAGACGAAAGTATACAGTGGCCTTGGGCCCATATG  279

Qy     57  GlyPheCysThrArgAspArgGluAspIleAsnSerLeuCysMetThrValValGlnAsnLeu  76
               ||:::||
Db    280  GGCTTCTGCTCAGTCCAAGAGACAACATCAACTCCCTGTGCCTGANNNTGGTGCAACGGCTG  339

Qy     77  MetGluArgGlnAsnAsnLeuSerThrTyrAspCysTyrIleGlyTyrArgLeuGluValGlyThrGluThr  96
               :::||||
Db    340  ATGGAGCGCCGCATACAGCTCCCATGGGACTCTGTGGGAGTAGGCACTGAGACC  399

Qy     97  IleIleLeuAspLysSerSerValLysSerThrAsnLeuMetGlnLeuPheGluGluSerGly  116
               :::||||
Db    400  ATCATTGACAAGTCCAAAGCTGTCAAAAACAGTGCCAACTCTTCCAGGANNCAGGC  459

Qy    117  AsnThrAspIleGluGlyGlyIleAspAspThrThrArgAsnAlaCysTyrArgGlyGlyThrAlaAlaVal  136
               :||||||
Db    460  AATACTGATATTGAGGGCGGCATAGNATACCACCAATGCCTGCTACGGTGGTGGTACTGCCTCCCTC  519

Qy    137  PheAsnAlaValAsnTrpIleGluSerSerTrpAsp  149
               |||||:
Db    520  TTCAATGCTGCCAACTGGATGGAGTCCAGTTCCTGGGATGGTCGTTATCCATGGTGGTC  579

Qy    149  ----------------------------------------------------  149
Db    580  TGTGGAGACATTGCCGTCTTATCCCAGTGTGTAAGTCTCGTCCCCAGAGGTGGGGGGCCGGAGCT  639

Qy    150  -------------------------------GlyLeuArgGlyThr  154
               ||||||:
Db    640  GTGGCTATGCTGATTGGGCCAAGGCCCCTTCGGCCGTTGAGGGGAACC  699

Qy    155  HisMetGlnHisAlaThrTyrTyrLysPheThrAspTyrSerProAspMetLeuSerGlyLeuTyrProIleIleVal  174
               |||:::||
Db    700  CATATGCAGCATGCGACTTATTACAAACCTACAAAATTTGGCCTGGAGTACCCAATAGTG  759

Qy    175  AspGlyGlyLeuSerIleGlnCysTyrArgLeuSerAlaLeuAspArgCysTyrTyrSerValTyr  194
               ||||||
Db    760  GATGGGGAAGCTTTCATCCAGTGCTACTTGCGGGCCCTTGGATGTTAACATCATAC  819

Qy    195  CysLysLysHisIleAlaGlnIleTrpGlnLeuAsnTrpLysSerAspPheThrLeuAsn  214
               |:::||
Db    820  CGTAAAAAAATCATGGAAATCAGTGGAAGCAAGCTGGACACCCTTCACCCTTGAC  879

Qy    215  AspAspGlyPheMetIlePheHisSerSerProTyrCysLysLysSerLeuAla  234
               ||:::||
Db    880  GATTACAGTACATGATCTTTCAATACACCCTTTTGCAAGAGTGGTTCCAGAAGTCTCTGGCT  939

Qy    235  ArgMetLeuLeuAsnAspPheLeuAsnAsnArgAspLysSerAsnSerIleTyrSer  254
               |||||:
Db    940  CGCCTGATGTTCAATGACTTCCTGTCAGCCGAGTGACACAAAACCAGCTTATATAAG  999

Qy    255  GlyIleLeuAlaPheGlyValLysAspValThrTyrPheAspArgAspValGlu  274
               |||:::||
Db   1000  GGGCTGGAGGCTTTCGGGGGGGTAAAGTGGAAGACTTACACCAACAAGGACCTGGAT  1059

Qy    275  LysAlaPheMetLysAlaSerSerGluLeuPheSerGlnIleLysThrLysAlaSerLeuLeu  294
               ||||||
Db   1060  AAAGCACTTCTAAAGGCCTCTCAGCACAGTTTCACACAAAAAACAGGGCTTCCTTTAC  1119

Qy    295  ValSerAsnGlnAsnGlyVaAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerVal  314
               :::||||
Db   1120  GTCTCCACGCACCAATGGGAACATGTACACCTCCATCCGTATACGGGTGTCGCTGGCTT  1179

Qy    315  LeuAlaGlnTyrSerProGlnIleLeuValAlaGlyTyrLysArgIleGlyValPheSerTyrGly  334
               ::||||
Db   1180  CTCTCCACGCACCAATGGGAACATGTACACCTCCATCCGTATACGGGTGTCGCTGGCTT
```

RESULT 10

```
BM474567

LOCUS           BM474567                 998 bp    mRNA    linear    EST    05-FEB-2002
DEFINITION      AGENCOURT_6489954 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5521787
                5', mRNA sequence.
ACCESSION       BM474567
VERSION         BM474567.1  GI:18523609
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 998)
  AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
  TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL       Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: ATCC
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Agencourt Bioscience Corporation
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM12187 row: p   column: 12
                High quality sequence stop: 759.
FEATURES             Location/Qualifiers
     source          1..998
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5521787"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT."
ORIGIN
```

Alignment Scores:

```
Pred. No.:                    4.12e-134
Score:                        1263.00
Percent Similarity:           83.33%
Best Local Similarity:        83.33%
Query Match:                  50.30%
DB:
```

```
Alignment Scores:
Length:            998
Matches:           250
Conservative:        0
Mismatches:          4
Indels:             46
Gaps:                1
```

US-10-622-516-2 (1-478) x BM474567 (1-998)

```
Qy      1  MetProGlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspValGlyIleVal  20
               |||||:
Db     30  ATGCCTGGATCACTTCCCTTTGAATGCGAAGGCTTGCGGCCAAAGATGTGGGAATTGTT  89

Qy     21  AlaLeuGluGlyIleGluTyrPheProSerGlnTyrValAspGlnAlaGluLeuGluLysTyrAsp  40
               ::||||
Db     90  GCCCTTGAGAGCTATTTTCCTTCCAATATGTTGATCAAGCAGAGAGTTGGAAAAATGAT  149
```

RESULT 11
BX419944

LOCUS   BX419944   1201 bp   mRNA   linear   EST 13-MAY-2003
DEFINITION BX419944 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE18 5-PRIME, mRNA sequence.
ACCESSION   BX419944
VERSION   BX419944.1 GI:30650862
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr; Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3098.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022BC09QP1cluster=3098.r. Contact :
            Feng Liang Email: fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022BC09QP1.
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE18"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a Not-I-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Alignment Scores:
    Pred. No.:
    Score: 1263.00   Matches: 256
    Percent Similarity: 81.90%   Conservative: 2
    Best Local Similarity: 81.27%   Mismatches: 12
    Query Match: 50.30%   Indels: 48
    DB: 13   Gaps: 3

US-10-622-516-2 (1-478) x BX419944 (1-1201)

RESULT 12

```
LOCUS       BU508388                 905 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10099759 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6502465
            5', mRNA sequence.
ACCESSION   BU508388
VERSION     BU508388.1  GI:22814621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14058  row: j  column: 02
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1..905
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6502465"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb."
```

ORIGIN

Alignment Scores:
```
Pred. No.:
Score:                   1255.50
Percent Similarity:      82.94%
Best Local Similarity:   82.61%
Query Match:             50.00%
DB:                      13
```

Length: 905
Matches: 247
Conservative: 1
Mismatches: 6
Indels: 45
Gaps: 2

```
Qy    159 AlaTyrAspPheTyrArgProAspMetLeuSerGluTyrProIleValAspGlyLysLeu 178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    690 GCCTATGGATTTTATACAAGCCTGATATGCTAATCCTATAGTAGGATAGAAACTC 749

Qy    179 SerIleGlnCysTyrLeuSerAlaLeuAlaLeuAspArgCysTyrSerValTyrCysLysLysIle 198
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    750 TCCATACAGTGCTACCCAGTGCATTAGACCCGTCTACTCTGTCTACTGCAAAAAGATC 809

Qy    199 HisAlaGlnTrpGlnLysGlyLysAsnAspLysAsp-PheThrLeuAsnAspPheGlyPh 218
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    810 CATGCCCAGTGCAGAAAGGGAAATGATAAAGATTTTACCTTGAATGATTTGGCTT 869

Qy    218 eMetIlePheHisSerProTyrCysLysLysLeuValGlnLysSerLeuAlaArgMetLeuLe 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 CATGATCTTTCACTCACCATATTGTAAACTGGTTCAGAAATCTTCAGCTCGGATGTTGCT 929

Qy    238 uAsnAspPheLeuAsnAspGlnAsnArgAsnSerIleTyrSerGlyLeuGluAl 258
          |||:||||||||||||||||||||||||||||||||||||||||||
Db    930 GAATGACTTCCTTAATG---MCAGATAGAGATAA-AATAGTAC-TATAGTGGGCCTGGGACCC 984

Qy    258 aPheGlyAspValLysGluAspThrTyrPheAspArgAsp 272
          |||:::|||||||::::
Db    985 -TTTGGGGGAKGTAAATWAGARGCAC---CTCTTGATAGAGAG 1023
```

RESULT 13

```
LOCUS       CA512392                 789 bp    mRNA    linear   EST 15-NOV-2002
DEFINITION  UI-R-FJ0-cpz-j-19-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
            UI-R-FJ0-cpz-j-19-0-UI 5', mRNA sequence.
ACCESSION   CA512392
VERSION     CA512392.1  GI:25003346
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

US-10-622-516-2 (1-478) x BU508388 (1-905)

```
Qy      9 AlaGluAlaCysTrpProLysProArgAspAspValGlyIleValAlaAlaLeuGluIleTyrPheProSer 28
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 GCAGAAGCTTGCTGGCCAAAAGATCGTGGAATTGTTGCCGCTTGAGATCTATTTTCCTTCT 71

Qy     29 GlnTyrValAspGlnAlaGlnLeuGluLysTyrAspGlyValValAsp-AlaGlyLysTyrTh 48
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 CAAATATGTTGATCAAGCAGAGTTGGAAAAATATGATGGTTGTAGATGGCTGGAAAGTATAC 131

Qy     48 rIleGlyLeuGlyGlnAlaGlnLeuGlyPheCysThrArgAspArgGluAspIleAsnSerIle 68
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 CAATTGGCTTGGGCGCCAGGCCCAAGATGGGCCTTCTGCACGATAGAGAGAGATATTAACTCT 191

Qy     88 yArgLeuGluValGlyTyrThrGluThrIleIleIleAspLysLysSerLysLysThrAsnLe 108
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 TTGCAGAGCTGTGGTTCAGAATCTTCATGAGAGAGAAATAACCTTTCCTATGATTGCATTGG 251

Qy    108 uMetGlnLeuPheGluGluSerGlyAsnThrAspIleGluGluGlyIleIleAspThrThrAsnAl 128
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 GCGGCTGGGAAGTTGGAACAGAGACAAATCATCGACAAAATCAAAGTCTGTGGAAGACTAATTT 311

Qy    148 pAsp----------------------------------- 149
          |||
Db    432 GGATGGCGACGGTATGCCCTGGTAGTTGCAGGAGATATTGCTGTATATGCCACAGGAAATGC 491

Qy    149 --------------------------------------- 149
Db    312 GATGCAGCTGTTTGAAGAGTCTGGGGAATACAGGAATCGACACAACTAATGC 371

Qy    150 -------GlyLeuArgGlyThrHisMetGlnHisAlaTyrAspPhePheTyrLysProAs 166
                 ||||||||||||||||||||||||||||||||||||||||||||||
Db    372 ATGCTAATGGAGGCACAGCGCTGCTGTCTTCAATGCCTGTTAACTGGATTGAGTCCAGCTCTG 431

Qy    166 pMetLeuSerGluTyrProIleValAspGlyLysLeuSerIleGlnCysTyrLeuSerAl 186
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 TTTTGAACAGAGGGCTTCGTCGGACACATATGCAACATGCCTATGATTTTTACAAGCCTGA 611

Qy    186 aLeuAlaAspArgCysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGlnLysGlyGlu 206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    612 TATGCTACTCTGAATAATCCTATAGTAGGATGGAAAAACTCCATACAGTGCTACCCTCAGTGC 671

Qy    206 yAsnAspLysAspPheThrLeuAsnAspPheGlyPhe-PheMetIlePheHisSerProTyrC 226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 ATTAGACACGCTGCTATTCTGTCTACTGCAAAAAGATCCCAATGCCCAGTGCCAGAAAGAGGG 731

Qy    226 ysLysLysLeuValGlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnA 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 AAATGGATAAGAGATTTTACTTTGAATGATTTGGGCTTCATGACATCTTTCACTCACCATATT 791

Qy    246 snArgAsnAspLysAsnSerIleTyrSerGlyLeu---GluAlaPheGlyAspVal 262
          |||||||||||||||||||||||||||||||||||:::
Db    792 GTAAACTGGNTCAGAAATCTTCAGCTCGGATGNTGCCGAAAGNGCCAGA 851

Db    852 ATAGAGATAAAAAAATAGTAATCTATAGTGGGGCCTGGGATGTT 904
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Ratus.
REFERENCE   1  (bases 1 to 789)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     9704477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. James Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 REVERSE.
FEATURES             Location/Qualifiers
     source          1..789
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-FJ0-cpz-j-19-0-UI"
                     /tissue_type="embryo"
                     /dev_stage="embryo"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-R-FJ0"
                     /note="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;
                     UI-R-FJ0 is a cDNA library containing the following
                     tissue(s): rat embryo. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CATCTCTACT. This library
                     was created for the University of Iowa Program for Rat
                     Gene Discovery and Mapping (Val Sheffield, Bento Soares
                     and Tom Casavant)."
ORIGIN

US-10-622-516-2 (1-478) x CAS12392 (1-789)

Alignment Scores:
Pred. No.:
Score:                  1.94e-132        Length:        789
Percent Similarity:     94.66%           Matches:       244
Best Local Similarity:  93.13%           Conservative:    4
Query Match:            49.66%           Mismatches:     14
DB:                     14               Indels:          0
                                         Gaps:            0

Qy  190 CysTyrSerValTyrCysLysLysIleHisAlaGlnTrpGlnLysGluGlyAsnAspLys 209
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TGCTATTCTGTCTACCGCAAAAAGATCCGGGCCCAGTGGCAGAAAGAGGGAAAGGATAAA  62

Qy  210 AspPheThrLeuAsnAspPheGlyPheMetIlePheHisSerProTyrCysGlyLeuVal 229
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 GATTTTACCCTGAATGATTTTGGCTTCATGATCTTTCACTCCCCATACTGTAAACTGGTG 122

Qy  230 GlnLysSerLeuAlaArgMetLeuLeuAsnAspPheLeuAsnAspGlnAsnArgAspLys 249
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 CAGAAATCTCTAGCTAGGATGCTCCTAGAACGATCAAAAACAGAGACAAA 182

Qy  250 AsnSerIleTyrSerGlyLeuGluAlaPheGlyAspValLysLeuGluAspThrTyrPhe 269
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183 AACAGTATTTACAGTGGGCTGGAAGCCTTTGGCGATGTGAAAATTAGAGGATACTTC 242

Qy  270 AspArgAspValGlyLysAlaPheMetLysAlaSerSerGluLeuPheSerGlyLysThr 289
        ||||||||||||||||||||::||||||||||||||||||||||||||||||||||||||
Db  243 GACAGAGATGTGGAAAAGGCATTTATGAAGGCTAGTGCTGAGCTATTCAACCAGAAAACA 302

Qy  290 LysAlaSerLeuLeuValSerAsnGlnAsnGlyAsnMetThrThrSerSerValTyrGly 309
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303 AAGGCAGCATCTTTGCTTGTATCGAATCAAAATGGAAACATGACAACATCCTCTGTATACGGT 362

Qy  310 SerLeuAlaSerValLeuAlaGlnTyrSerProGlnLnLeuLeuAlaGlyLysIleGly 329
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 TCCCTTGCTTCTGTTCTGGCACAGTACTCACCTCAACAGTTGGCCCCGGGAAGAGGATTGGA 422

Qy  330 ValPheSerTyrGlySerGlyLeuAlaAlaThrLeuTyrSerLeuLeuValThrGlnAsp 349
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  423 GTGTTCTCTTACGGTTCTGGCTTGGCAGCTACTCTGTACTCCCTTAAAGTCACACAAGAT 482

Qy  350 AlaThrProGlySerLeuAlaLeuAspLysIleThrAlaSerLeuCysAspLysLysSerArg 369
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483 GCCACACACCAGGATCTGCTCTTGACAAAATAACAGCAAGTTTATGTGACCTTAAGTCAAGG 542

Qy  370 LeuAspSerArgThrGlyValAlaProAspValPheAlaLeuGluAsnMetLysLeuArgGlu 389
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  543 CTTGACTCAAGAACGGGTGGCACCAGACGTCTTTGCTGAAAACATGAAGCTGAGAGAG 602

Qy  390 AspThrHisIleSerLeuValAsnTyrIleProGlnGlySerIleGluAspSerLeuPheGlyGly 409
        :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  603 GACACACATCACTTAGCCAACTATATATTCCCGCAGGGTTCAATAGAATTCACTCTTCCGAAGGA 662

Qy  410 ThrTrpTyrLeuValArgValAspGluLysSerArgArgThrTyrAlaArgArgProThr 429
        |||||||||||||||||||||||||||||||||||||||||||||||||:::
Db  663 ACATGGTATCTAGTCAGAGTGGATGAAAAGACACAGAGAACTTACGCCCGGCGTCCCTCC 722

Qy  430 ProAsnAspAspThrLeuAspGluGlyValGlyLeuValHisSerAsnIleAlaThrGlu 449
        ::::|||||||||||||||||||||||||||||||||||:::|||||||||
Db  723 ACANNNATGACACAGTTGGGATGAAGGAGTGGGACTTGTGCATTCANACACAGCTACAGAG 782

Qy  450 HisIle 451
        ||||||
Db  783 CATATT 788


RESULT 14
CD354310
LOCUS       CD354310              779 bp    mRNA    linear    EST 15-JUL-2003
DEFINITION  UI-M-GM0-cgcc-j-20-0-UI.r1 NIH_BMAP_GM0 Mus musculus cDNA clone
            IMAGE:30360907 5', mRNA sequence.
ACCESSION   CD354310
VERSION     CD354310.1  GI:31146811
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 779)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/mousef1.html
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..779
```

ORIGIN

```
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30360907"
/tissue_type="whole brain"
/dev_stage="1, 5 and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_GM0"
/note="Organ: Brain; Vector: pYX- Asc; Site_1: EcoR I;
Site_2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was size
selected according to mRNA size fraction, ligated with EcoR
I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The size fraction was
sequence located between the Not I site and the polyA tail
is CGAACTGAAT.This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
```

US-10-622-516-2 (1-478) x CD354310 (1-779)

Alignment Scores:

Pred. No.:		
Score:	7.04e-129	Length:	779
Percent Similarity:	1216.00	Matches:	239
Best Local Similarity:	94.59%	Conservative:	6
Query Match:	92.28%	Mismatches:	14
DB:	48.43%	Indels:	0
	14	Gaps:	0

```
Qy  219 MetIlePheHisSerProThrTyrCysLysGlyLeuValGlnLysSerLeuAlaArgMetLeuLeu 238
            ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db    3 ATGATCTTTCACTCACCATATTGTAAACTGGTGCAGAAAATCCTAGCTCGGATGTTCCTG 62

Qy  239 AsnAspPheLeuAsnAspGlnAsnArgAspGlyAsnSerIleTyrSerGlyLeuGluAla 258
            ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db   63 AATGACTTTCTTAATGATCAAAACAGAGATAAAAACAGTATTTACAGTGGACTGGAAGCC 122

Qy  259 PheGlyAspValLysLeuLeuGluArgThrTyrPheAspArgAspValGlyLysAlaPheMet 278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 TTTGGGGACGTTAAATTAGAAGATACTTACTTTGACAGAGATGTCGGGAAAGCATTTATG 182

Qy  279 LysAlaSerSerGluLeuPheSerGlnLysThrLysAlaSerLeuLeuValSerAsnGln 298
            ||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||
Db  183 AAGGCTAGTTCTGAGCTATTCAGCCAGAAAACAAAGGCCTCTTTGCTTGTGTCTAATCAG 242

Qy  299 AsnGlyAsnMetTyrThrSerSerValTyrGlySerLeuAlaSerValLeuAlaGlnTyr 318
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 AATGGAAATATGTACACATCCTCTGTTTATGGTTCCCTGGCTTCTGTCCTGGCACAGTAC 302

Qy  319 SerProGlnGlnLeuAlaGlyLysArgIleGlyValPheSerTyrGlySerGlyLeuAla 338
            ||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db  303 TCACCTCAGCAGTTGGCAGGGAAGAGGGTTGAGTGTTTCTTACGGTTCTGGCTTGGGCTT 362

Qy  339 AlaThrLeuTyrSerLeuGlyValThrGlnAspAlaThrProGlySerAlaLeuAspLys 358
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363 GCCACGCTGTACTCCCTTAAAGTCACACAAGACCCACACCAGGATCTCCCCTTGATAAA 422

Qy  359 IleThrAlaSerLeuCysLysSerArgLeuAspSerArgThrGlyValAlaPro 378
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  423 ATAACAGCAAGTTATGTGACCCTTAAATCAAGCCTTGACTCGAGAACTGGTGTGGCACCG 482

Qy  379 AspValPheAlaGluAsnMetLysLeuLeuArgGluAspThrHisHisLeuValAsnTyrIle 398
            ||||||||||||||:|||||||||||||||||||||||||||:|||||||||||||||
Db  483 GATGTCTTTGCTGAAAACATGAAGCTACTCAGAGAGGACACACATCACTTAGCCAACTATATT 542

Qy  399 ProGlnGlySerIleAspSerLeuPheGluGlyThrTrpTyrLeuValArgValAspGlu 418
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 15

BU149049

LOCUS   BU149049   967 bp   mRNA   linear   EST 03-SEP-2002
DEFINITION   AGENCOURT_8743315 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6392260
        5', mRNA sequence.
ACCESSION   BU149049
VERSION   BU149049.1   GI:22662581
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 967)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
        Email: cgapbs-r@mail.nih.gov
        Tissue Procurement: Susan L. Sullivan, PhD.
        cDNA Library Preparation: ResGen, Invitrogen Corp
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLAM13881 row: n column: 05
        High quality sequence stop: 599.
FEATURES   Location/Qualifiers
    source   1..967
        /organism="Mus musculus"
        /mol_type="mRNA"
        /db_xref="taxon:10090"
        /clone="IMAGE:6392260"
        /clone="DH10B (phage-resistant)"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_129"
        /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1;
        Site_1: EcoRV, Site_2: NotI; Cloned unidirectionally.
        Primer: Oligo dT. Average insert size 2.2 kb. Constructed
        by ResGen, Invitrogen Corp. Note: this is a NIH_MGC
        Library."

ORIGIN

```
Db  543 CCCCAGTGTTCAATAGACTCACTCCTTTGAAGGAAGGTGGTATCTGGTCAGAGTGGATGA 602
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  419 LysHisArgArgThrTyrAlaArgArgProThrProAsnAspAspArgThrLeuAspGly 438
            ||||:|||||||||||||||||||||||||||||:||||||:||||||||||||||||
Db  603 AAACACAGAAGGACTTACGCGCCGGCGCCCCCTTCACANATGACCACAGTTTGGATGAAGGA 662

Qy  439 ValGlyLeuValHisSerAsnIleAlaThrGluHisSerIleProSerProAlaLysLysVal 458
            ::|||||||||||||||||||||::||||||||||||||||||||||||||||||||||
Db  663 ATGGGGCCTCGTGCATAGTAACACACAGCAAGACAGAGCATATTCCAAGCCCTGCTAAGAAACT 722

Qy  459 ProArgLeuProAlaThrAlaAlaGluProGluAlaValAlaValIleSerArgAsnGlyGlu 477
            ||||||||||||||||||||||||:||:|||||:|||||||||||||||||||||||||
Db  723 GCAAGACTCCCTGCAACCTCGGGNCGAAATCGAATCAGCTGTCATCAGTAGCTAACGGGGAG 779
```

US-10-622-516-2 (1-478) x BU149049 (1-967)

Alignment Scores:

Pred. No.:		
Score:	4.95e-128	Length:	967
Percent Similarity:	1210.00	Matches:	249
Best Local Similarity:	80.82%	Conservative:	8
Query Match:	78.30%	Mismatches:	16
DB:	48.19%	Indels:	47
	13	Gaps:	1

```
Qy    3 GlySerLeuProLeuAsnAlaGluAlaCysTrpProLysAspAspValGlyIleValAlaLeu 22
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22 GGGTCACTTCCTTTGAAATGCAGAGGCTTGCCTGCCCAAAAGAGATGTGGGAATCGTTCGCCCTT 81

Qy   23 GluIleTyrPheProSerGlnTyrValAspGlnAlaGluGluGlyLeuGlyLysTyrArgAspGlyVal 42
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    82  GAAATCTACTTTCCTTCTCAATATGTCGATCAAGCTGAGTTGGAAAAATACGATGGGTGTA 141
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        AspAlaGlyLysTyrThrIleGlyTyrLeuGlyLeuGlyAlaLysMetGlyPheCysThrAspArg 62
Qy    43  GATGCTGGAAAAGTATACCATCGGCCCTGGGCTCAGGGCCAGGATGGGTTCTGCACGGATCGT 142

Qy    63  GGAGACATCAACTCTCTTTGCCTGACTGGTGGTTCAGAAACTGATGGAGAGACAATAGCCTT 202
Db        GluAspIleAsnSerLeuCysMetThrValValGlnAsnLeuuMetGluArgAsnAsnLeu 82
Qy   202  GAAGACATCAACTCTCTTTGCCTGACTGGTGGTTCAGAAACTGATGGAGAGACAATAGCCTT 261

Qy    83  TCCTTATGATTGCATTGGGCGGCGGCCTAGAAGTTGGAACAGAGACAAATCCAAA 321
Db        SerTyrAspCysIleGlyArgLeuGluValGlyThrGluThrIleIleleAspLysLysSerLys 102
Qy   262  TCCTTATGATTGCATTGGGCGGCGGCCTAGAAGTTGGAACAGAGACAAATCCAAA 321

Qy   103  TCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTTCTGGGAAATACAGATATAGAAGGA 381
Db        SerValLysSerAsnLeuMetGlnLeuPheGluGluLysSerGlyAsnThrAspIleGluGly 122
Qy   322  TCAGTGAAGTCTAATTTGATGCAGCTGTTTGAGGAGTTCTGGGAAATACAGATATAGAAGGA 381

Qy   123  ATAGATACAACCAAATGCATACCTATGGGGGGCACAGCGTGCAGTCTTCAAATGCCGTGAACTGG 441
Db        IleAspThrThrAsnAlaCysTyrGlyGlyThrAlaAlaValPheAsnAlaValAsnTrp 142
Qy   382  ATAGATACAACCAAATGCATACCTATGGGGGCACAGCGTGCAGTCTTCAAATGCCGTGAACTGG 441

Qy   143  GTCGGAATCCAGCTCTTGGGGATGGACGATATGCTTCTGGTAGTTGCAGGACACATTGCTATA 501
Db        IleGluSerSerSerTrpAspGly----------------------------------- 150
Qy   442  GTCGGAATCCAGCTCTTGGGGATGGACGATATGCTTCTGGTAGTTGCAGGACACATTGCTATA 501

Qy   150  -----------------------------------------------------------  150
Db

Qy   151  -----------------------------LeuArgGlyThrrHisMetGlnHisAlaTyr 160
Db        TATGCCACAGGAAAATGCCAGACCTACAGGTGGAGTTGGAGCGTGGCCCCTGCTAATTGGG 561
Qy   502  TATGCCACAGGAAAATGCCAGACCTACAGGTGGAGTTGGAGCGTGGCCCCTGCTAATTGGG 561

Qy   161  GACTTTTTACAGGCCTGACATGCTCTCCGAGTACCCGTGGTGTGGACGGAAAAGTCTCCATA 681
Db        AspPheTyrLysProCysAspMetLeuSerGluTyrProIleValAspGlyLysLeuSerIle 180
Qy   622  GACTTTTTACAGGCCTGACATGCTCTCCGAGTACCCGTGGTGTGGACGGAAAAGTCTCCATA 681

Qy   181  CAGTGCTACCTCAGCCCTGGACCGCTGCTATTCGTCTACCCGCAAAAGATCCGGTGCC 741
Db        GlnCysTyrLeuSerAlaLeuAspArgCysTyrSerValTyrCysLysLysIleHisAla 200
Qy   682  CAGTGCTACCTCAGCCCTGGACCGCTGCTATTCGTCTACCCGCAAAAGATCCGGTGCC 741

Qy   201  CAGTGGCGAGAAAGAGGGAAAAAGGATPAAGA-TTTACCCTGAATGAAGTNTTCATGATC 800
Db        GlnTrpGlnLysGlyLysAsnAspLysAspPheThrLeuAsnAspPheGlyPheMetIle 220
Qy   742  CAGTGGCGAGAAAGAGGGAAAAAGGATPAAGA-TTTACCCTGAATGATNTTCATGATC 800

Qy   221  TTTCACTCACCATATTGTACTGGTGCAGAAATCTTCTAGCTCCGGATGTCCCTGAATGAC 860
Db        PheHisSerProTyrCysLysLysGlnValGlnLysSerLeuAlaArgMetLeuLeuAsnAsp 240
Qy   801  TTTCACTCACCATATTGTAACTGGTGCAGAAATCTTCTAGCTCCGGATGTCCCTGAATGAC 860

Qy   241  TTNTCCTTATGATCAAAAACAGAGATAA-AACAGTATTTACTGGACTGGAAGCCTTTTGG 919
Db        PheLeuAsnAspGlnAsnArgAspLysAsnSerIleTyrSerGlyLeuGluAla-PheGl 260
Qy   861  TTNTCCTTATGATCAAAAACAGAGATAA-AACAGTATTTACTGGACTGGAAGCCTTTTGG 919

Qy   260  yAsp-ValLysLeuGluAspThrTyr-PheAspArgAspValGluLys 275
Db        GGACGNNTAATTAGNAGATACTTACTTTTGACCGAGATGTAAAAATAAG 967
Qy   920  GGACGNNTAATTAGNAGATACTTACTTTTGACCGAGATGTAAAAATAAG 967
```